

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 07:13:37 ; Search time 486 Seconds
(without alignments)
10560.297 Million cell updates/sec

Title: US-10-024-197-24
Perfect score: 2279
Sequence: 1 agctaaagcaggtacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|-------------|--------------------|
| 1 | 1659.4 | 72.8 | 1661 | 20 AAX26682 | Complete coding se |
| 2 | 1659.4 | 72.8 | 1661 | 21 AAA48444 | Human lysosomal gl |
| 3 | 1654.2 | 72.6 | 1741 | 22 AAF28293 | Unmodified gcc DNA |
| 4 | 1654.2 | 72.6 | 1741 | 22 AAF28294 | Modified gcc DNA. |
| 5 | 1649.8 | 72.4 | 1661 | 9 AAN80113 | Encodes recombinan |
| 6 | 1601.6 | 70.3 | 1642 | 18 AAT71753 | Human glucocerebro |
| 7 | 1587 | 69.6 | 1611 | 22 AAD06389 | Human glucocerebro |
| 8 | 1547.8 | 67.9 | 1551 | 22 AAD06388 | Human glucocerebro |
| 9 | 1544 | 67.7 | 1593 | 24 AAH99925 | Nucleotide sequenc |

| | | | | | |
|----|--------|------|-------|-------------|--------------------|
| 10 | 1543 | 67.7 | 1551 | 22 AAS10863 | Human cDNA encodin |
| 11 | 1543 | 67.7 | 1551 | 24 ABK40235 | DNA encoding human |
| 12 | 1529.6 | 67.1 | 1592 | 24 AAH99924 | Nucleotide sequenc |
| 13 | 667 | 29.3 | 7620 | 14 AAQ39286 | Glucocerebrosidase |
| 14 | 639.4 | 28.1 | 5769 | 24 ABL67250 | Thyroid cancer rel |
| 15 | 363 | 15.9 | 417 | 21 AAC00312 | Human secreted pro |
| 16 | 209.6 | 9.2 | 229 | 16 AAT20128 | Human gene signatu |
| 17 | 191.2 | 8.4 | 834 | 22 AAL20417 | Human breast cance |
| 18 | 175.2 | 7.7 | 384 | 23 ABV37314 | Human prostate exp |
| 19 | 134.8 | 5.9 | 246 | 16 AAT25845 | Human gene signatu |
| 20 | 130.4 | 5.7 | 138 | 24 ABL38256 | Human colon tumour |
| 21 | 106 | 4.7 | 243 | 23 ABV07380 | Human prostate exp |
| 22 | 92.4 | 4.1 | 4861 | 23 ABL16351 | Drosophila melanog |
| 23 | 89.4 | 3.9 | 358 | 14 AAQ48243 | IVS2+1 mutant gluc |
| 24 | 78.6 | 3.4 | 9517 | 23 ABL16350 | Drosophila melanog |
| 25 | 77 | 3.4 | 89 | 14 AAQ39304 | Mutant glucocerebr |
| 26 | 75 | 3.3 | 75 | 14 AAQ39287 | Wild-type glucocer |
| 27 | 68 | 3.0 | 412 | 22 AAF67412 | Novel human polynu |
| 28 | 60 | 2.6 | 60 | 24 ABN32368 | Human spliced tran |
| 29 | 59.6 | 2.6 | 159 | 22 AAL11519 | Human breast cance |
| 30 | 53 | 2.3 | 1528 | 22 AAF76268 | Trichoderma harzia |
| 31 | 47.4 | 2.1 | 65 | 24 ABN53669 | Mouse spliced tran |
| 32 | 45.4 | 2.0 | 55 | 24 AAH99930 | Oligonucleotide 2 |
| 33 | 40 | 1.8 | 40 | 22 AAF28290 | Unmodified gcc con |
| 34 | 39.4 | 1.7 | 41 | 22 AAF28288 | Unmodified gcc con |
| 35 | 38.4 | 1.7 | 477 | 24 ABL80290 | Human ovarian canc |
| 36 | 38.4 | 1.7 | 2367 | 22 AAQ94257 | Human full-length |
| 37 | 38.4 | 1.7 | 2410 | 23 AAS92255 | DNA encoding novel |
| 38 | 38.4 | 1.7 | 2442 | 22 AAQ94333 | Human full-length |
| 39 | 38.4 | 1.7 | 2673 | 21 AAC76488 | Human ORFX ORF2043 |
| 40 | 38.4 | 1.7 | 3919 | 24 ABK54223 | Human HEAT-3 DNA. |
| 41 | 38.4 | 1.7 | 3950 | 24 AAD27270 | Human transporter |
| 42 | 38.4 | 1.7 | 9561 | 23 AAS94283 | DNA encoding novel |
| 43 | 38.2 | 1.7 | 390 | 21 AAC04310 | Human secreted pro |
| 44 | 37.8 | 1.7 | 399 | 22 AAH34387 | Human colon cancer |
| 45 | 37.6 | 1.6 | 19286 | 22 AAL03002 | Human reproductive |

ALIGNMENTS

RESULT 1

AAX26682

ID AAX26682 standard; DNA; 1661 BP.

XX AAX26682;

AC AAX26682;

XX AAX26682;

DT 17-JUN-1999 (first entry)

XX Complete coding sequence for lysosomal glucocerebrosidase.

DE Human lysosomal glucocerebrosidase; glycosylated; Gaucher's disease; ss.

XX Human lysosomal glucocerebrosidase; glycosylated; Gaucher's disease; ss.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Eliason WK, Ginns EJ, LaMarca ME, Martin B, Maysak KA;
XX WPI; 1999-203905/17.
DR P-PSDB; AAY01595.
DR
XX
PT New compositions containing glucocerebrosidase - comprising
PT glycosylated recombinantly-produced human glucocerebrosidase, used
PT for treating Gaucher's disease
XX
PS Disclosure; Fig 1A-D; 10pp; English.
CC The present sequence encodes a human lysosomal glucocerebrosidase.
CC The specification describes a composition containing glycosylated
CC recombinantly produced human glucocerebrosidase which can be used
CC for treating Gaucher's disease.
XX
SQ Sequence 1661 BP; 374 A; 496 C; 422 G; 369 T; 0 other;
Query Match 72.8%; Score 1659.4; DB 20; Length 1661;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 108 GGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCAT 167
Db 1 GGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCAT 60
QY 168 GGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTCTGGGCATCAGGTGCCG 227
Db 61 GGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTCTGGGCATCAGGTGCCG 120
QY 228 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATA 287
Db 121 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATA 180
QY 288 CTGTGACTCCTTTGACCCCCCGACCTTCTCTGCTTGGTACCTTCAGCCGCTATGAGAG 347
Db 181 CTGTGACTCCTTTGACCCCCCGACCTTCTCTGCTTGGTACCTTCAGCCGCTATGAGAG 240
QY 348 TACACGCAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGG 407
Db 241 TACACGCAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGG 300
QY 408 CACAGGCCTGTACTGACCCCTGCAGCAGAACAGAGTTCCAGAAAGTGAAGGGATTGG 467
Db 301 CACAGGCCTGTACTGACCCCTGCAGCAGAACAGAGTTCCAGAAAGTGAAGGGATTGG 360
QY 468 AGGGCCATGACAGATGCTGCTGCTCAACATCCTTGCCTGTACCCCTGCCCAAAA 527
Db 361 AGGGCCATGACAGATGCTGCTGCTCAACATCCTTGCCTGTACCCCTGCCCAAAA 420
QY 528 TTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 587
Db 421 TTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 480
QY 588 CATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATT 647
Db 481 CATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATT 540
QY 648 CCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 707
Db 541 CCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 600
QY 708 CCGAGCCCTGAGTGGCCCGACGCTCCGTTTCACTCCTTGCAGCCCTGGACATCACC 767
Db 601 CCGAGCCCTGAGTGGCCCGACGCTCCGTTTCACTCCTTGCAGCCCTGGACATCACC 660
QY 768 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCG 827
Db 661 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCG 720
QY 828 AGACATCTACACACAGACCTGGGCCAGATACTTTGTGAAGTTCCTGGATGCCTATGCTGA 887
XX

Db 721 AGACATCTACCACAGACCTGGGCCAGATACTTTGTGAAGTTCTCGATGCCCTATGCTGA 780
QY 888 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAAAATGAGCCTTCTGCTGGCTGTTGAG 947
Db 781 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAAAATGAGCCTTCTGCTGGCTGTTGAG 840
QY 948 TGGATACCCCTTCCAGTGCCTGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCCCG 1007
Db 841 TGGATACCCCTTCCAGTGCCTGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCCCG 900
QY 1008 TGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACATGTCCGCTACTCATGCTGA 1067
Db 901 TGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACATGTCCGCTACTCATGCTGA 960
QY 1068 TGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1127
Db 961 TGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1128 TAAATATGTTTCATGTCATGCTGTACATTTGCTACCTGGACTTCTGGCTCCAGCCAAAGC 1187
Db 1021 TAAATATGTTTCATGTCATGCTGTACATTTGCTACCTGGACTTCTGCTCCAGCCAAAGC 1080
QY 1188 CACCCTAGGGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTTGCTCAGAGGCCCTG 1247
Db 1081 CACCCTAGGGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTTGCTCAGAGGCCCTG 1140
QY 1248 TGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGATCGAGGGATGCA 1307
Db 1141 TGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGATCGAGGGATGCA 1200
QY 1308 GTACAGCCACAGCATCATCACGAACCTCTGTACCATGTGGTGGCTGGACCGACTGGAA 1367
Db 1201 GTACAGCCACAGCATCATCACGAACCTCTGTACCATGTGGTGGCTGGACCGACTGGAA 1260
QY 1368 CCTTGCCCTGAACCCGAGGAGGAGGCCCAATGGGTGCGTAACCTTTGTCACAGTCCCAT 1427
Db 1261 CCTTGCCCTGAACCCGAGGAGGAGGCCCAATGGGTGCGTAACCTTTGTCACAGTCCCAT 1320
QY 1428 CATGTAGACATCACCAAGGACACAGTTTACAAACAGCCCATGTTCTTACACACTTGGCCA 1487
Db 1321 CATGTAGACATCACCAAGGACACAGTTTACAAACAGCCCATGTTCTTACACACTTGGCCA 1380
QY 1488 CTTAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGCTGGTGGCTGCTGCTGCTGCTG 1547
Db 1381 CTTAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGCTGGTGGCTGCTGCTGCTGCTG 1440
QY 1548 CGACCTGGACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1607
Db 1441 CGACCTGGACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 1608 CCGTCTCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGCTGCTGCTGCTGCTGCTG 1667
Db 1501 CCGTCTCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGCTGCTGCTGCTGCTGCTG 1560
QY 1668 AATCTACCTGGCTTACCTTACACCTTACACCTTACCTTACCTTACCTTACCTTACCTTAC 1727
Db 1561 AATCTACCTGGCTTACCTTACACCTTACACCTTACCTTACCTTACCTTACCTTACCTTAC 1620
QY 1728 CTCAAGGAGGACCTGGGCTCAGCCTGGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1768
Db 1621 CTCAAGGAGGACCTGGGCTCAGCCTGGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1661
RESULT 2
AAA48444
ID AAA48444 standard; DNA; 1661 BP.
XX
AC AAA48444;
XX
DT 08-SEP-2000 (first entry)
XX
DE Human lysosomal glycoprotein glucocerebrosidase (GCS) coding sequence.
XX

KW Human; lysosomal glycoprotein glucocerebrosidase; Gaucher's disease;
KW GCS; beta-D-glucosyl-N-acylsphingosine glycohydrolase; neuroprotective;
KW hepatotropic; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 59..1609
FT /*tag= a
FT /EC_number= 3.2.1.45
FT /product= "GCS"
FT sig_peptide 59..115
FT /*tag= b
FT mat_peptide 116..1606
FT /*tag= c

XX US6074864-A.

XX 13-JUN-2000.

XX 13-JAN-1994; 94US-0186256.

XX 23-DEC-1987; 87US-0137796.

XX 05-FEB-1990; 90US-0474307.

XX 06-AUG-1992; 92US-0925333.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Ginns EI, Maysak KA, Eliason WK, Lamarca ME, Martin B;

XX WPI; 2000-430396/37.

XX P-PSDB; AAY99599.

XX New glycosylated human glucocerebrosidase useful for treating Gaucher's
PT disease, is produced by transfecting Spodoptera frugiperda cells with
PT baculovirus containing cDNA encoding glucocerebrosidase -

XX Claim 4; Fig 1; 12pp; English.

XX The present sequence is the coding sequence for human lysosomal
CC glycoprotein glucocerebrosidase (GCS, beta-D-glucosyl-N-acylsphingosine
CC glycohydrolase). Mutation or deficiency of the GCS gene leads to
CC Gaucher's disease. In the present invention, a Spodoptera frugiperda
CC insect cell infected with a recombinant baculovirus vector containing a
CC the present sequence has been developed. This cell line is useful for
CC large-scale production of GCS, which could be used to treat Gaucher's
CC disease via gene therapy.

XX Sequence 1661 BP; 374 A; 496 C; 422 G; 369 T; 0 other;

Query Match 72.8%; Score 1659.4; DB 21; Length 1661;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 1660; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 108 GGAGTTTCAAGTCTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCAT 167

Db 1 GGAGTTTCAAGTCTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCAT 60

QY 168 GGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTCTGGGCATCAGGTGCCCG 227

Db 61 GGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTCTGGGCATCAGGTGCCCG 120

QY 228 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGTGTCTGCAATGCCACATA 287

Db 121 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGTGTCTGCAATGCCACATA 180

QY 288 CTGTGACTCCTTTGACCCCGCCGACCTTCCTGCCCCCTGGTACCTTCACCCGCTATGAGAG 347

Db 181 CTGTGACTCCTTTGACCCCGCCGACCTTCCTGCCCCCTGGTACCTTCACCCGCTATGAGAG 240

QY 348 TACACGAGTGGCGGAGGATGGAGTGAGTATGGGCGCCATCCAGGTAATCACACGGG 407

Db 241 TACACGAGTGGCGGAGGATGGAGTGAGTATGGGCGCCATCCAGGTAATCACACGGG 300

QY 408 CACAGGCCTGCTACTGACCCTGCAGCCAGAACAGAGTTCAGAAAAGTCAAGGGATTGG 467

Db 301 CACAGGCCTGCTACTGACCCTGCAGCCAGAACAGAGTTCAGAAAAGTCAAGGGATTGG 360

QY 468 AGGGCCATGACAGATGCTGCTCTCTCAACACATCCTTGCCCTGTACCCCTGCCCAAAA 527

Db 361 AGGGCCATGACAGATGCTGCTCTCTCAACACATCCTTGCCCTGTACCCCTGCCCAAAA 420

QY 528 TTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 587

Db 421 TTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 480

QY 588 CATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACACCCCTGATGATT 647

Db 481 CATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACACCCCTGATGATT 540

QY 648 CCAGTTGCACAACTTACGCTTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 707

Db 541 CCAGTTGCACAACTTACGCTTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 600

QY 708 CCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTTCACCTCTGCCAGCCCTGGACATCAC 767

Db 601 CCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTTCACCTCTGCCAGCCCTGGACATCAC 660

QY 768 CACTTGGCTCAAGACCATGGAGCGGTGAATGGGAGGGGTCACTCAAGGGACAGCCCG 827

Db 661 CACTTGGCTCAAGACCATGGAGCGGTGAATGGGAGGGGTCACTCAAGGGACAGCCCG 720

QY 828 AGACATCTACCACAGACCTGGGCCAGATACTTTGTGAAGTTCCTGGATGCCCTATGCTGA 887

Db 721 AGACATCTACCACAGACCTGGGCCAGATACTTTGTGAAGTTCCTGGATGCCCTATGCTGA 780

QY 888 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGGCTGTTGAG 947

Db 781 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGGCTGTTGAG 840

QY 948 TGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCG 1007

Db 841 TGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCG 900

QY 1008 TGACCTAGTCTCTACCTCGCCAAACAGTACTCACACAAATGTCCGCTACTCATGCTGGA 1067

Db 901 TGACCTAGTCTCTACCTCGCCAAACAGTACTCACACAAATGTCCGCTACTCATGCTGGA 960

QY 1068 TGACCAACGCTTGTGCTGCCCTCACCTGGCAAGGTGTTACTGACAGACCCAGAACG 1127

Db 961 TGACCAACGCTTGTGCTGCCCTCACCTGGCAAGGTGTTACTGACAGACCCAGAACG 1020

QY 1128 TAAATATGTTTCAATGGCATGCTGTACATTTGGTACCTGGACTTTCTGGCTCCAGCCAAAG 1187

Db 1021 TAAATATGTTTCAATGGCATGCTGTACATTTGGTACCTGGACTTTCTGGCTCCAGCCAAAG 1080

QY 1188 CACCCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCTG 1247

Db 1081 CACCCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCTG 1140

QY 1248 TGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGCTAGGCTCCTGGGATCGAGGGATGCA 1307

Db 1141 TGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGCTAGGCTCCTGGGATCGAGGGATGCA 1200

QY 1308 GTACAGCCACAGCATCATCACGAACCTCTCTGTACCATGTTGGTCTGGCTGGACCGACTGGA 1367

Db 1201 GTACAGCCACAGCATCATCACGAACCTCTCTGTACCATGTTGGTCTGGCTGGACCGACTGGA 1260

QY 1368 CCTTGGCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCTGCTAACCTTTGTCAGAGTCCC 1427

Db 1261 CCTTGGCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCTGCTAACCTTTGTCAGAGTCCC 1320

QY 1428 CATTTGATAGCATCACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGCCA 1487

Db 1321 CATTTGATAGCATCACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGCCA 1380

QY 1488 CTTAGCAAGTTTCATTCCTGAGGGCTCCCGAGAGAGTGGGGCTGGTGGCCAGTCAGAAGAA 1547
|||||
Db 1381 CTTAGCAAGTTTCATTCCTGAGGGCTCCCGAGAGAGTGGGGCTGGTGGCCAGTCAGAAGAA 1440
|||||
QY 1548 CGACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTTGTTGGTGGTAA 1607
|||||
Db 1441 CGACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTTGTTGGTGGTAA 1500
|||||
QY 1608 CCGCTCCTCTAAGGATGTCCTCTTACCATCAAGGATCCTGCTGTTGGGCTCCTGGAGAC 1667
|||||
Db 1501 CCGCTCCTCTAAGGATGTCCTCTTACCATCAAGGATCCTGCTGTTGGGCTCCTGGAGAC 1560
|||||
QY 1668 AATCTCACCTGGCTACTCCATTACACACCTAGCTGTGGCGTGGCCAGTGGAGCAGATA 1727
|||||
Db 1561 AATCTCACCTGGCTACTCCATTACACACCTAGCTGTGGCGTGGCCAGTGGAGCAGATA 1620
|||||
QY 1728 CTCAAGGAGGACCTGGGCTCAGCCTGGGCTGAGGATCAAGGAGCA 1768
|||||
Db 1621 CTCAAGGAGGACCTGGGCTCAGCCTGGGCTGAGGATCAAGGAGCA 1661
|||||

RESULT 3
AAF28293
ID AAF28293 standard; DNA; 1741 BP.
AC AAF28293;
XX
DT 04-APR-2001 (first entry)
DE Unmodified Gcc DNA.
XX
KW Glucocerebrosidase; Gcc; splice; gene therapy; enzyme therapy;
KW Gaucher; ds.
XX
OS Unidentified.
XX
PN CA2272055-A1.
XX
PD 02-DEC-2000.
XX
PF 02-JUN-1999; 99CA-2272055.
XX
PR 02-JUN-1999; 99CA-2272055.
XX
PA (HSCR-) HSC RES & DEV LP.
PI Mahuran DJ, Callahan JW, Clarke JTR;
XX
DR WPI; 2001-160192/17.
XX

PT New glucocerebrosidase (Gcc) DNA molecule modified in at least one
PT nucleotide, useful in gene therapy or enzyme replacement therapy,
PT particularly in treating Gaucher disease -
XX
PS Disclosure; Fig 4; 37pp; English.
XX
CC The present invention relates to glucocerebrosidase (Gcc) DNA
CC with a modification in at least one nucleotide that disrupts a
CC splicing consensus sequence and prevents splicing of mRNA produced
CC from the DNA molecule, while preserving the ability of the DNA to
CC express active Gcc. The glucocerebrosidase (Gcc) DNA molecule and
CC the vector comprising the DNA are useful in gene therapy or enzyme
CC replacement therapy, particularly in treating Gaucher disease. The
CC cells transfected or transduced by the vectors are useful as research
XX tools.
SQ Sequence 1741 BP; 393 A; 519 C; 441 G; 387 T; 1 other;
Query Match 72.6%; Score 1654.2; DB 22; Length 1741;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 78 CTTCTCTTCAATGACCTTGAGGGGATGGAGTTTCAAGTCTTCCAGAGAGGAATG 137
|||||

Db 9 CTTAGCTTGAATTAAGAGGCGGACGCCATGGAGTTTCAAGTCTTCCAGAGAGGAATG 68
|||||
QY 138 TCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTCACAGGATTTGCTTCTACT 197
|||||
Db 69 TCCCAAGCCTTTGAGTAGAGTCTCCATCATGGCTGGCAGCCTCACAGGTTTGTCTTCTACT 128
|||||
QY 198 TCAGGCAGTGTCTGTTGGCATCAGGTGCCCGCCCTGCATCCCTAAAGCTTCGGCTACAG 257
|||||
Db 129 TCAGGCAGTGTCTGTTGGCATCAGGTGCCCGCCCTGCATCCCTAAAGCTTCGGCTACAG 188
|||||
QY 258 TCGGCTGGTGTGTCTGCAATGCCACATACTGTGACTCTTTGACCCCGGACCTTTCC 317
|||||
Db 189 TCGGCTGGTGTGTCTGCAATGCCACATACTGTGACTCTTTGACCCCGGACCTTTCC 248
|||||
QY 318 TCGCCTTGGTACCTTCAGCGCTATGAGAGTACACGAGTGGCGACGGATGGAGCTGAG 377
|||||
Db 249 TCGCCTGGGAACATTTTCCCGCTATGAGAGTACACGAGTGGCGACGGATGGAGCTGAG 308
|||||
QY 378 TATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTTCGAGCCAGA 437
|||||
Db 309 TATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTTCGAGCCAGA 368
|||||
QY 438 ACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGSCCATGACAGATGCTGCTCTCAA 497
|||||
Db 369 ACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGSCCATGACAGATGCTGCTCTCAA 428
|||||
QY 498 CATCCTTGGCCTGTACCCCTGCCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGA 557
|||||
Db 429 CATCCTTGGCCTGTACCCCTGCCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGA 488
|||||
QY 558 AGGAATCGGATATAACATCATCCGGGTACCCATGGCAGCTGTGACTTCTCATCCGCAC 617
|||||
Db 489 AGGAATCGGATATAACATCATCCGGGTACCCATGGCAGCTGTGACTTCTCATCCGCAC 548
|||||
QY 618 CTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGCTCCAGAGGA 677
|||||
Db 549 CTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGCTCCAGAGGA 608
|||||
QY 678 AGATACCAAGCTCAAGTACCCCTGATTCACCGAGCCTGCAGTTGGCCAGCGTCCCGT 737
|||||
Db 609 AGATACCAAGCTCAAGTACCCCTGATTCACCGAGCCTGCAGTTGGCCAGCGTCCCGT 668
|||||
QY 738 TTCACCTCCTTGGCAGCCTCGGACATCACCCACCTGGCTCAAGACCAATGGAGCGGTGAA 797
|||||
Db 669 TTCACCTCCTTGGCAGCCTCGGACATCACCCACCTGGCTCAAGACCAATGGAGCGGTGAA 728
|||||
QY 798 TGGGAAGGGGTCACTCAAGGGACAGCCCGGAGAGATCTACCACAGACCTGGGCCAGATA 857
|||||
Db 729 TGGGAAGGGGTCACTCAAGGGACAGCCCGGAGAGATCTACCACAGACCTGGGCCAGATA 788
|||||
QY 858 CTTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTGGCAGTGACAGC 917
|||||
Db 789 CTTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTGGCAGTGACAGC 848
|||||
QY 918 TGAAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 977
|||||
Db 849 TGAAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 908
|||||
QY 978 CCGTGAACATCAGCGAGACTTCATTGCCCGTACCTAGGTCTCTACCCCTCGCAACAGTAC 1037
|||||
Db 909 CCGTGAACATCAGCGAGACTTCATTGCCCGTACCTAGGTCTCTACCCCTCGCAACAGTAC 968
|||||
QY 1038 TCACCACAAATGTCCCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCCCGGCTGGGC 1097
|||||
Db 969 TCACCACAAATGTCCCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCCCGGCTGGGC 1028
|||||
QY 1098 AAAGGTGGTACTGACAGACCCAGAGAGCAGCTAAATATGTTTCATGGCATGCTGTACATTG 1157
|||||
Db 1029 AAAGGTGGTACTGACAGACCCAGAGAGCAGCTAAATATGTTTCATGGCATGCTGTACATTG 1088
|||||
QY 1158 GTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTAGGGGAGACACACCGCTGTTCCTCC 1217
|||||

QY 858 CTTTGTGAAGTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTTGGCAGTGACAGC 917
|||||
Db 789 CTTTGTGAAGTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTTGGCAGTGACAGC 848
QY 918 TGAATAATGAGCCTTCTCTGGCTGCTGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGCTTCC 977
|||||
Db 849 TGAATAATGAGCCTTCTCTGGCTGCTGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGCTTCC 908
QY 978 CCCTGAACATCAGGAGACTTCAATTGCCGTGACCTAGGTCTTACCTCGCCAAACAGTAC 1037
|||||
Db 909 CCCTGAACATCAGGAGACTTCAATTGCCGTGACCTAGGTCTTACCTCGCCAAACAGTAC 968
QY 1038 TCACCACAATGTCCGCTTACTCATGCTGGATGACCAACGCTTCTGCTGCCCTGAGC 1097
|||||
Db 969 TCACCACAATGTCCGCTTACTCATGCTGGATGACCAACGCTTCTGCTGCCCTGAGC 1028
QY 1098 AAAGGTGGTACTGACAGACCCAGAACGACGCTAAATATGTTTCATGSCATTTGCTGATCTG 1157
|||||
Db 1029 AAAGGTGGTACTGACAGACCCAGAACGACGCTAAATATGTTTCATGSCATTTGCTGATCTG 1088
QY 1158 GTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCCGCTGTTCCC 1217
|||||
Db 1089 GTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCCGCTGTTCCC 1148
QY 1218 CAACACCATGCTTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCCGCTGTTCCC 1277
|||||
Db 1149 CAACACCATGCTTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCCGCTGTTCCC 1208
QY 1278 GCGGCTAGGCTCTGGGATCGAGGGATGAGTACAGCCACAGCATCATCACGAACCTCCT 1337
|||||
Db 1209 GCGGCTAGGCTCTGGGATCGAGGGATGAGTACAGCCACAGCATCATCACGAACCTCCT 1268
QY 1338 GTACCATGTGGTGGCTGGACCGACTGGAACCTTGCCCTGAACCCGAGGAGGACCCAA 1397
|||||
Db 1269 GTACCATGTGGTGGCTGGACCGACTGGAACCTTGCCCTGAACCCGAGGAGGACCCAA 1328
QY 1398 TTGGGTGCGTAACCTTGTGACAGTCCCATCATTTGATGACATCACCAAGGACACGTTTAA 1457
|||||
Db 1329 TTGGGTGCGTAACCTTGTGACAGTCCCATCATTTGATGACATCACCAAGGACACGTTTAA 1388
QY 1458 CAACACGCCCCATGTTCTACCACTTGGCCACTTCCAGCAAGTTCTTCTGAGGGCTCCCA 1517
|||||
Db 1389 CAACACGCCCCATGTTCTACCACTTGGCCACTTCCAGCAAGTTCTTCTGAGGGCTCCCA 1448
QY 1518 GAGAGTGGGCTGGTGGCTGAGTCCAGAACGACGACCTGGACGCGAGTGGCATGATCATCC 1577
|||||
Db 1449 GAGAGTGGGCTGGTGGCTGAGTCCAGAACGACGACCTGGACGCGAGTGGCATGATCATCC 1508
QY 1578 CGATGGCTCTGCTGTGTGGTGGTGGCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1637
|||||
Db 1509 CGATGGCTCTGCTGTGTGGTGGTGGCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1568
QY 1638 CAAGGATCCTGCTGTGGGCTTCTCTGGAGACAACTCTACCTGGCTACTCCATTACACCTA 1697
|||||
Db 1569 CAAGGATCCTGCTGTGGGCTTCTCTGGAGACAACTCTACCTGGCTACTCCATTACACCTA 1628
QY 1698 CCTGTGGCTCGCCAGTGGAGCAGATGAGGAGGACGCTGAGGCTCAGCCTGGGCA 1757
|||||
Db 1629 CCTGTGGCTCGCCAGTGGAGCAGATGAGGAGGACGCTGAGGCTCAGCCTGGGCA 1688
QY 1758 TTAAAGGGACAGAGTCCAGC 1776
|||||
Db 1689 TTAAAGGGACAGAGTCCAGC 1707

RESULT 5
AAN80113
ID AAN80113 standard; cDNA; 1661 BP.
XX
AC AAN80113;
XX
DT 17-DEC-2001 (updated)
DT 12-OCT-1990 (first entry)

XX Encodes recombinant human lysosomal glucocerebrosidase.
DE lysosomal glycoprotein glucocerebrosidase ; Gaucher's disease; ss.
KW synthetic.
OS
FH Key Location/Qualifiers
FT CDS 59..1607
FT /*tag= a
FT /product=recombinant human lysosomal glucocerebrosidase
XX USN7137796-N.
PN
XX
PD 02-AUG-1988.
XX
PF 23-DEC-1987; 87US-0192026.
XX
PR 23-DEC-1987; 87US-0137796.
XX
PA (USSH) US DEPT HEALTH AND HUMAN SERVICES.
XX
PI Ginns EI;
XX
DR WPI; 1988-264274/37.
DR P-PSDB; AAP80109.
XX
PT Recombinant glucocerebrosidase -
PT useful for treating Gaucher's disease
XX
PS Disclosure; ; 2pp; English.
XX
CC The sequence is used for construction of the baculovirus derived
CC vector PAC373/GC. It differs from cDNA sequences published
CC previously, for example leu (489) was previously given as Pro and
CC Arg (514) was previously given as His.
CC See also AAN80112.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ Sequence 1661 BP; 374 A; 496 C; 422 G; 369 T; 0 other;

Query Match 72.4%; Score 1649.8; DB 9; Length 1661;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1654; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 108 GGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAACCCCTTTGAGTAGGGTAAGCATCAT 167
|||||
Db 1 GGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAACCCCTTTGAGTAGGGTAAGCATCAT 60
QY 168 GGCTGGCAGCCTCACAGGATGCTTCTACTTACAGGAGTGTGCTGGGCATCAGGTGCCCG 227
|||||
Db 61 GGCTGGCAGCCTCACAGGATGCTTCTACTTACAGGAGTGTGCTGGGCATCAGGTGCCCG 120
QY 228 CCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACATA 287
|||||
Db 121 CCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACATA 180
QY 288 CTGTGACTCCTTTGACCCCGGACCTTTCTCCTGCCCTTGGTACCTTACAGCCGCTATGAGAG 347
|||||
Db 181 CTGTGACTCCTTTGACCCCGGACCTTTCTCCTGCCCTTGGTACCTTACAGCCGCTATGAGAG 240
QY 348 TACAGCAGTGGCGACGAGTGGAGCTGAGTATGGGCCCATCCAGGCTAAATCACACGGG 407
|||||
Db 241 TACAGCAGTGGCGACGAGTGGAGCTGAGTATGGGCCCATCCAGGCTAAATCACACGGG 300
QY 408 CACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTCCAGAAAGTGAAGGATTTGG 467
|||||
Db 301 CACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTCCAGAAAGTGAAGGATTTGG 360
QY 468 AGGGCCCATGACAGATGCTGTGCTCTCAACATCCTTGCCCTGTCAACCCCTGCCCAAAA 527

Db 361 AGGGCCATGACAGATGCTGCTCTCAACATCCTTGCCCTGTCAACCCCTGCCAAA 420
QY 528 TTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 587
Db 421 TTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACC 480
QY 588 CATGGCCAGCTGTGACTTCTCCATCCGGACCTACACCTATGCAGACACCCCTGATGATT 647
Db 481 CATGGCCAGCTGTGACTTCTCCATCCGGACCTACACCTATGCAGACACCCCTGATGATT 540
QY 648 CCAGTTGCACAACCTTCAAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 707
Db 541 CCAGTTGCACAACCTTCAAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 600
QY 708 CCGAGCCCTGCAGTTGGCCCGCAGCGTCCCGTTTCACTCTTCCAGCCCTGGACATCAC 767
Db 601 CCGAGCCCTGCAGTTGGCCCGCAGCGTCCCGTTTCACTCTTCCAGCCCTGGACATCAC 660
QY 768 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGG 827
Db 661 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGG 720
QY 828 AGACATCTACACCAAGCTGGGCGGAGATCTTGTGAAGTTCTTGGATGCTATGCTGA 887
Db 721 AGACATCTACACCAAGCTGGGCGGAGATCTTGTGAAGTTCTTGGATGCTATGCTGA 780
QY 888 GCACAAAGTTACAGTTCTGGGCGAGTGACAGCTGAAATAGCCCTTCTGCTGGGCTGTTGAG 947
Db 781 GCACAAAGTTACAGTTCTGGGCGAGTGACAGCTGAAATAGCCCTTCTGCTGGGCTGTTGAG 840
QY 948 TGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAACATCAGCGAGACTTTCATTGCCCG 1007
Db 841 TGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAACATCAGCGAGACTTTCATTGCCCG 900
QY 1008 TGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACATGTCCGCTACTCATGCTGGA 1067
Db 901 TGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACATGTCCGCTACTCATGCTGGA 960
QY 1068 TGACCAACCGCTTGTCTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGC 1127
Db 961 TGACCAACCGCTTGTCTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGC 1020
QY 1128 TAAATATGTTTATGGATTTGCTGTACATTTGGTACCTGGACTTCTTGGCTCCAGCCAAAGC 1187
Db 1021 TAAATATGTTTATGGATTTGCTGTACATTTGGTACCTGGACTTCTTGGCTCCAGCCAAAGC 1080
QY 1188 CACCTAGGGGAGACACACCGCTGTTCCTCCCAACACCATGCTTTCCTCAGAGGCTG 1247
Db 1081 CACCTAGGGGAGACACACCGCTGTTCCTCCCAACACCATGCTTTCCTCAGAGGCTG 1140
QY 1248 TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATGCA 1307
Db 1141 TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATGCA 1200
QY 1308 GTACAGCCACAGCATCATCAGCAACCTCCTGTACCATGTGGTGGCTGGACCGACATGGA 1367
Db 1201 GTACAGCCACAGCATCATCAGCAACCTCCTGTACCATGTGGTGGCTGGACCGACTGGA 1260
QY 1368 CCTTCCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCGTAACCTTGTGACAGTCCCAT 1427
Db 1261 CCTTCCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCGTAACCTTGTGACAGTCCCAT 1320
QY 1428 CATTTAGACATCACCAAGGACAGCTTTTACAAACAGCCCATGTTCTACCACTTGGCCA 1487
Db 1321 CATTTAGACATCACCAAGGACAGCTTTTACAAACAGCCCATGTTCTACCACTTGGCCA 1380
QY 1488 CTTACAGCAAGTTTATCTCCTGAGGGCTCCAGAGAGTGGGCTGGTGGCAGTCAAGAA 1547
Db 1381 GTTCAGCAAGTTTATCTCCTGAGGGCTCCAGAGAGTGGGCTGGTGGCAGTCAAGAA 1440
QY 1548 CGACCTGGACGAGTGGCACTGATGATCCCGATGGCTCTGCTGTTGTTGCTGCTAAA 1607

Db 1441 CGACCTGGACGAGTGGCATTTGATGCATCCCGATGGCTCTGCTGTTGTTGCTGCTAAA 1500
QY 1608 CCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCCTGCTGTGGGCTTCTGGAGAC 1667
Db 1501 CCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCCTGCTGTGGGCTTCTGGAGAC 1560
QY 1668 AATCTCACCTGGCTACTCTCCATTCACACCTACCTGTGGCTGCCAGTGTGAGCAGATA 1727
Db 1561 AATCTCACCTGGCTACTCTCCATTCACACCTACCTGTGGCTGCCAGTGTGAGCAGATA 1620
QY 1728 CTCAAGGAGGCACTGGGCTCAGCCTGGGCATTAAGGGACA 1768
Db 1621 CTCAAGGAGGCACTGGGCTCAGCCTGGGCATTAAGGGACA 1661

RESULT 6
AAT71753
ID AAT71753 standard; cDNA; 1642 BP.
XX AAT71753;
AC AAT71753;
XX 01-OCT-1997 (first entry)
DT
XX Human glucocerebrosidase-FLAG cDNA construct.
DE
XX Lysosomal enzyme; lysosome; transgenic plant; glucocerebrosidase;
KW enzyme replacement therapy; Gaucher disease; ss.
KW
XX Chimaeric Homo sapiens;
OS Chimaeric synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1642
FT /*tag= a
FT /transl_except= (pos:265..267, aa:Glu)

XX WO9710353-A1.
XX 20-MAR-1997.
XX 13-SEP-1996; 96WO-US14730.
XX 14-SEP-1995; 95US-0003737.
XX (CROP-) CROPTech DEV CORP.
PA (VIRG) VIRGINIA TECH INTELLECTUAL PTY INC.
XX
PI Cramer CL, Oishi KK, Radin DN, Weissenborn DL;
XX WPI; 1997-202248/18.
DR P-PSDB; AAW18237.
XX
XX Production of enzymatically active (modified) lysosomal enzyme in
transgenic plants - useful in treatment of lysosomal storage
disorders
XX
XX Example 1; Page 61-62; 111pp; English.

CC A cDNA construct (AAT71753) codes for a fusion protein (AAW18237)
comprising human glucocerebrosidase (hGC) and a C-terminal FLAG
epitope (see also AAW18236). The hGC sequence was obtd. by PCR
amplification using primers (see also AAT71755-56) designed to
incorporate restriction sites to facilitate cloning and ligation
of the FLAG coding sequence. The construct, placed under control
of the inducible Mega promoter (see also AAT71752), can drive
expression of hGC in transgenic (esp. tobacco) plants. The plant
expression system provides for post-translational modification and
processing to produce enzymatically active hGC, which can be
harvested for use in Gaucher disease enzyme replacement therapy.

SQ Sequence 1642 BP; 368 A; 492 C; 415 G; 367 T; 0 other;
Query Match 70.3%; Score 1601.6; DB 18; Length 1642;

30-JUN-2000; 2000US-0215430.
(AVIG-) AVIGEN INC.
(CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
PI Podsakoff G, Watson G, Couto LB, Yang B;
XX WPI; 2001-343814/36.
DR P-PSDB; AAE02446.
XX
PT Use of recombinant adeno-associated virus, comprising gene encoding a
PT protein defective or missing in lysosomal storage disease, in the
PT manufacture of a medicament for treating the lysosomal storage disease
PT -
XX
PS Example 3A; Page 91-94; 97pp; English.
XX
CC The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful
CC in the manufacture of a medicament for treating lysosomal storage
CC disease e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome)
CC is due to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS)
CC which aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in
CC gene therapy. The present sequence is human glucocerebrosidase (GC)
CC cDNA. Mutations in GC gene leads to Gaucher's disease, a lysosomal
CC storage disease. This sequence is used in AAV constructs.
XX
SQ Sequence 1611 BP; 359 A; 487 C; 404 G; 361 T; 0 other;

Query Match 69.6%; Score 1587; DB 22; Length 1611;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1596; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 106 ATGGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCCTTTGAGTAGGGTAAGCATC 165
DB 1 ATGGAGTTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCCTTTGAATGACCCCTGAGGGG 60

QY 166 ATGGCTGGCAGCCTCAGAGGATTGCTTCTACTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 225
DB 61 ATGGCTGGCAGCCTCAGAGGTTTGTCTTCTACTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 120

QY 226 CGCCCTGCATCCCTAAAGCTTCCGCTACAGCTCGGTGGTGTGTGTGCTGCAATGCCACA 285
DB 121 CGCCCTGCATCCCTAAAGCTTCCGCTACAGCTCGGTGGTGTGTGTGCTGCAATGCCACA 180

QY 286 TACTGTGACTCCTTTGACCCCTCCGACCTTCTGCTGCTTGGTACCTTTCAGCCGCTATGAG 345
DB 181 TACTGTGACTCCTTTGACCCCTCCGACCTTCTGCTGCTTGGTACCTTTCAGCCGCTATGAG 240

QY 346 AGTACAGCAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGC 405
DB 241 AGTACAGCAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGC 300

QY 406 GGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAAAGTGAAGGGATT 465
DB 301 GGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAAAGTGAAGGGATT 360

QY 466 GGAGGGCCCATGACAGATGCTGCTGCTCTCAACATCCTTGCCCTGTCACCCCTGCCCAA 525
DB 361 GGAGGGCCCATGACAGATGCTGCTGCTCTCAACATCCTTGCCCTGTCACCCCTGCCCAA 420

QY 526 AATTGCTACTTAAATCGTACTTCTGAAAGAGGAATCGGATATATACATCATCCGGSTA 585
DB 421 AATTGCTACTTAAATCGTACTTCTGAAAGAGGAATCGGATATATACATCATCCGGSTA 480

QY 586 CCCATGGCCAGCTGTGACTTCTCCATCCGCAACCTACCTATGACAGACCCCTGATGAT 645
DB 481 CCCATGGCCAGCTGTGACTTCTCCATCCGCAACCTACCTATGACAGACCCCTGATGAT 540

QY 646 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAGGAATACCAAGCTCAAGATACCCCTGATT 705
DB 541 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAGGAATACCAAGCTCAAGATACCCCTGATT 600

QY 706 CACCGAGCCCTGCAAGTTGGCCAGCGTCCCGTTTCACTCCTTTGCCAGCCCTTGACATCA 765
DB 601 CACCGAGCCCTGCAAGTTGGCCAGCGTCCCGTTTCACTCCTTTGCCAGCCCTTGACATCA 660

QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGACAGCCC 825
DB 661 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGACAGCCC 720

QY 826 GGAGACATCTACACACAGACCTGGGCCAGATACTTTGTGAAGTTCTTGATGCTATGCT 885
DB 721 GGAGACATCTACACACAGACCTGGGCCAGATACTTTGTGAAGTTCTTGATGCTATGCT 780

QY 886 GAGCACAAAGTTACAGTTCTGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 945
DB 781 GAGCACAAAGTTACAGTTCTGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 840

QY 946 AGTGATACCCCTTCCAGTGGCTGGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCC 1005
DB 841 AGTGATACCCCTTCCAGTGGCTGGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCC 900

QY 1006 CGTGACCTAGGCTCCTACCCCTGCCAACAGTACTCACCACAAATGTCCGCTTACTCATGCT 1065
DB 901 CGTGACCTAGGCTCCTACCCCTGCCAACAGTACTCACCACAAATGTCCGCTTACTCATGCT 960

QY 1066 GATGACCAACGCTTGTGCTGCCCCACTGGSCAAAGGTGGTACTGACAGACCCAGAGCA 1125
DB 961 GATGACCAACGCTTGTGCTGCCCCACTGGSCAAAGGTGGTACTGACAGACCCAGAGCA 1020

QY 1126 GCTAAATATGTTTCATGGCAATGCTGTACATGGTACCTGGACTTTCTGGCTCCAGCCAAA 1185
DB 1021 GCTAAATATGTTTCATGGCAATGCTGTACATGGTACCTGGACTTTCTGGCTCCAGCCAAA 1080

QY 1186 GCCACCCCTAGGGGAGACACACACCGCCTGTTCCCAACACCATGCTCTTTGCCCTCAGAGGCC 1245
DB 1081 GCCACCCCTAGGGGAGACACACACCGCCTGTTCCCAACACCATGCTCTTTGCCCTCAGAGGCC 1140

QY 1246 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTTAGGCTCCTGGGATCGAGGGATG 1305
DB 1141 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTTAGGCTCCTGGGATCGAGGGATG 1200

QY 1306 CAGTACAGCCACAGCATCATCACGAACCTCCTGTACCATGTGGTGGCTGGACCCGACTGG 1365
DB 1201 CAGTACAGCCACAGCATCATCACGAACCTCCTGTACCATGTGGTGGCTGGACCCGACTGG 1260

QY 1366 AACCTTGGCTGAACCCCGAAGGAGGACCAATTTGGTGGCTAACTTTGTGACAGTCCC 1425
DB 1261 AACCTTGGCTGAACCCCGAAGGAGGACCAATTTGGTGGCTAACTTTGTGACAGTCCC 1320

QY 1426 ATCATTTGATAGACATCACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGC 1485
DB 1321 ATCATTTGATAGACATCACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGC 1380

QY 1486 CACTTCAGCAAGTTCAATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAGAG 1545
DB 1381 CACTTCAGCAAGTTCAATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAGAG 1440

QY 1546 AACGACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTGCTGTTGTTGGTCTGCTA 1605
DB 1441 AACGACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTGCTGTTGTTGGTCTGCTA 1500

QY 1606 AACCGCTCCTCTAAAGGATGCTCCTCTTACCATCAAGGATCCTGCTGTGGGCTTCCCTGGAG 1665
DB 1501 AACCGCTCCTCTAAAGGATGCTCCTCTTACCATCAAGGATCCTGCTGTGGGCTTCCCTGGAG 1560

QY 1666 ACAATCTACCTGGCTTACTCCATTCACACCTACCTACCTGCGGCTGCCAGTGA 1716
DB 1561 ACAATCTACCTGGCTTACTCCATTCACACCTACCTACCTGCGGCTGCCAGTGA 1611

RESULT 8
AAD06388
ID AAD06388 standard; cdna; 1551 BP.

XX
AC
XX
DT
XX
DE
XX
KW
KW
KW
KW
XX
OS
XX
FH
FT
FT
FT
FT
FT
FT
FT
FT
XX
PN
XX
PD
XX
PF
XX
PR
PR
XX
PA
PA
XX
PI
XX
DR
DR
XX
PT
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
SQ

AAD06388;
10-AUG-2001 (first entry)
Human glucocerebrosidase (GC) cDNA #1.
Human; adeno-associated viral expression vector; AAV; gene therapy;
lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG;
glucocerebrosidase; GC; Gaucher's disease; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 1..1551
/tag= a
/product= "Human glucocerebrosidase (GC)"
sig_peptide 1..57
/tag= b
mat_peptide 58..1548
/tag= c
/product= "Human mature glucocerebrosidase (GC)"
WO200136603-A2.
25-MAY-2001.
17-NOV-2000; 2000WO-US31688.
17-NOV-1999; 99US-0166097.
30-JUN-2000; 2000US-0215430.
(AVIG-) AVIGEN INC.
(CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
Podsakoff G, Watson G, Couto LB, Yang B;
WPI; 2001-343814/36.
P-PSDB; AAE02445.
Use of recombinant adeno-associated virus, comprising gene encoding a
protein defective or missing in lysosomal storage disease, in the
manufacture of a medicament for treating the lysosomal storage disease
.
Example 3A; Page 87-89; 97pp; English.
The present invention relates to recombinant adeno-associated virus (AAV)
expression vectors and virions, which include genes coding for enzymes
defective or missing in lysosomal storage disease (LSD). AAV is useful
in the manufacture of a medicament for treating lysosomal storage
disease e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome)
is due to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS)
which aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in
gene therapy. The present sequence is human glucocerebrosidase (GC)
cDNA. Mutations in GC gene leads to Gaucher's disease, a lysosomal
storage disease. This sequence is used in AAV constructs.
Sequence 1551 BP; 344 A; 474 C; 387 G; 346 T; 0 other;
Query Match 67.9%; Score 1547.8; DB 22; Length 1551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTCTGGGTCATCAGGTGCC 225
|||||
DB 1 ATGGCTGGCAGCCCTCACAGGTTTGTCTTCTACTTTCAGGCAGTGTCTGGGTCATCAGGTGCC 60
QY 226 CGCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACA 285
|||||
DB 61 CGCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACA 120

QY 286 TACTGTGACTCCTTTGACCCCGGACCTTTCTCCTGCCCTTGGTACCTTTCAGCCGCTATGAG 345
|||||
DB 121 TACTGTGACTCCTTTGACCCCGGACCTTTCTCCTGCCCTTGGTACCTTTCAGCCGCTATGAG 180
QY 346 AGTACACGACGTGGCGGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACG 405
|||||
DB 181 AGTACACGACGTGGCGGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACG 240
QY 406 GGACAGGCTGCTACTGACCTTGCAGCCGACGAGTATGGGGCCCATCCAGGCTAATCACACG 465
|||||
DB 241 GGACAGGCTGCTACTGACCTTGCAGCCGACGAGTATGGGGCCCATCCAGGCTAATCACACG 300
QY 466 GGAGGGCCCATGACAGATGCTGCTCTCAACATCCTTGCCTTGCCTGTACCCCTGCCCAA 525
|||||
DB 301 GGAGGGCCCATGACAGATGCTGCTCTCAACATCCTTGCCTTGCCTGTACCCCTGCCCAA 360
QY 526 AATTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAAATCATCCGGGTA 585
|||||
DB 361 AATTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAAATCATCCGGGTA 420
QY 586 CCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGAT 645
|||||
DB 421 CCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGCACAACTTCAAGCTCCGAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
|||||
DB 481 TTCCAGTTGCACAACTTCAAGCTCCGAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCCCTGGACATCA 765
|||||
DB 541 CACGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCCC 825
|||||
DB 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCCC 660
QY 826 GGAGACATCTACCACGACCTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCTGCTATGCT 885
|||||
DB 661 GGAGACATCTACCACGACCTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCTGCTATGCT 720
QY 886 GAGCACAAAGTTACAGTTCTGGCGAGTGACAGCTGAAATGAGGCTTCTGCTGGGCTGTG 945
|||||
DB 721 GAGCACAAAGTTACAGTTCTGGCGAGTGACAGCTGAAATGAGGCTTCTGCTGGGCTGTG 780
QY 946 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAAACATCAGCGAGCTTCAATGCC 1005
|||||
DB 781 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAAACATCAGCGAGCTTCAATGCC 840
QY 1006 CGTGACCTAGGCTCTACCTCGGCAACAGTACTCACCACCAATGTCCGCTTACTCATGCTG 1065
|||||
DB 841 CGTGACCTAGGCTCTACCTCGGCAACAGTACTCACCACCAATGTCCGCTTACTCATGCTG 900
QY 1066 GATGACCAAGCTTGTGCTGCTGCCACTGGGCAAGGTGTTACTGACAGACCCAGAGCA 1125
|||||
DB 901 GATGACCAAGCTTGTGCTGCTGCCACTGGGCAAGGTGTTACTGACAGACCCAGAGCA 960
QY 1126 GCTAATATGTTTCAATGGCATGCTGTTACATTTGGTACCTGCTGCTTCCAGCCAAA 1185
|||||
DB 961 GCTAATATGTTTCAATGGCATGCTGTTACATTTGGTACCTGCTGCTTCCAGCCAAA 1020
QY 1186 GCCACCTTAGGGGAGACACACCGCTGTTCCCAACACCAATGCTTTCCTCAGAGGCC 1245
|||||
DB 1021 GCCACCTTAGGGGAGACACACCGCTGTTCCCAACACCAATGCTTTCCTCAGAGGCC 1080
QY 1246 TGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1305
|||||
DB 1081 TGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCACGACCTCCTGTACCATGTGTGGCTGGACCGACTGG 1365
|||||
DB 1141 CAGTACAGCCACAGCATCATCACGACCTCCTGTACCATGTGTGGCTGGACCGACTGG 1200
QY 1366 AACCTTGGCTTGAACCCCGAAGGAGGAGGCCAAATTTGGGTGCGTAACCTTGTGCGACAGTCCC 1425

QY 526 AATTGTGCTACTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAAATCAATCATCCGGGTA 585
Db 361 AATTGTGCTACTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAAATCAATCATCCGGGTA 420
QY 586 CCATGGCCAGGTGTGACTTCTCCATCCGACCTACACCTATGAGAGACACCCCTGATGAT 645
Db 421 CCATGGCCAGGTGTGACTTCTCCATCCGACCTACACCTATGAGAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAGGATACCAAGCTCAAGATACCCCTGATT 705
Db 481 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAGGATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACCGAGCCCTGAGTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCCCTCGACATCA 765
Db 541 CACCGAGCACCTGAGTTGGCCAGCGTCCCGTTTCACTCCTTGCAGCCCTCGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGATGGGAAGGGTCACTCAAGGACAGCCC 825
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGATGGGAAGGGTCACTCAAGGACAGCCC 660
QY 826 GGAGACATCTACACACAGACCTGGGCCAGATACCTTGTGAAGTTCTTGGATGCTATGCT 885
Db 661 GGAGACATCTACACACAGACCTGGGCCAGATACCTTGTGAAGTTCTTGGATGCTATGCT 720
QY 886 GAGCACAAAGTTACAGTTCTGGCAGTGACAGTGAATAGCCTTCTGCTGGGCTGTTG 945
Db 721 GAGCACAAAGTTACAGTTCTGGCAGTGACAGTGAATAGCCTTCTGCTGGGCTGTTG 780
QY 946 AGTGGATACCCCTTCCAGTGGCTGGGCTTCAACCATGACGAGACATTCATTGCC 1005
Db 781 AGTGGATACCCCTTCCAGTGGCTGGGCTTCAACCATGACGAGACATTCATTGCC 840
QY 1006 CGTGACCTAGGTCCTACCCCTCGCCAAACAGTACTCAACCAATGTCCGCTTACTCATGCTG 1065
Db 841 CGTGACCTAGGTCCTACCCCTCGCCAAACAGTACTCAACCAATGTCCGCTTACTCATGCTG 900
QY 1066 GATGACCAACGCTGTGCTGCTGCCCCACTGGGCAAGGTGTTACTGACAGACCCAGAGCA 1125
Db 901 GATGACCAACGCTGTGCTGCTGCCCCACTGGGCAAGGTGTTACTGACAGACCCAGAGCA 960
QY 1126 GCTAAATATGTTCATGGCATGCTGTACATTTGGTACCTGGACTTCTTGGCTCCAGCCAAA 1185
Db 961 GCTAAATATGTTCATGGCATGCTGTACATTTGGTACCTGGACTTCTTGGCTCCAGCCAAA 1020
QY 1186 GCCACCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTGGCTCAGAGGCC 1245
Db 1021 GCCACCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTGGCTCAGAGGCC 1080
QY 1246 TGTGTGGGCTCAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCAGGGATG 1305
Db 1081 TGTGTGGGCTCAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCAGGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCAGCAACCTTCTGTACCATGTGTGGCTGGACCGACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGCAACCTTCTGTACCATGTGTGGCTGGACCGACTGG 1200
QY 1366 AACCTTGGCCTGAACCCCGAGGAGGACCCAAATGGGTGCGTAACCTTTGTCGACAGTCCC 1425
Db 1201 AACCTTGGCCTGAACCCCGAGGAGGACCCAAATGGGTGCGTAACCTTTGTCGACAGTCCC 1260
QY 1426 ATCATTTGTAGCATCACCAGGACACGCTTTTACAAACAGCCCATGTTCTACCACTTGGC 1485
Db 1261 ATCATTTGTAGCATCACCAGGACACGCTTTTACAAACAGCCCATGTTCTACCACTTGGC 1320
QY 1486 CACTTCAGCAAGTTTCAATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCCAGTCAAGAAG 1545
Db 1321 CATTTTCAGCAAGTTTCAATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCCAGTCAAGAAG 1380
QY 1546 AACGACCTGGACGAGTGGCCTGATGATCCCGATGGCTCTGCTGTTGTTGGTGGTGTGCTA 1605
Db 1381 AACGACCTGGACGAGTGGCCTGATGATCCCGATGGCTCTGCTGTTGTTGGTGGTGTGCTA 1440

QY 1606 AACCGCTCCTCTAAGGATGTGCCTCTTACCATCAAGGATCCTGTGTGGGCTTCTGGAG 1665
Db 1441 AACCGCTCCTCTAAGGATGTGCCTCTTACCATCAAGGATCCTGTGTGGGCTTCTGGAG 1500
QY 1666 ACAATCTCACCTGGCTACTCCATTCACTACACCTACCTGTGGGCTGCCAGTGA 1716
Db 1501 ACAATCTCACCTGGCTACTCCATTCACTACACCTACCTGTGGGCTGCCAGTGA 1551
RESULT 11
ABK40235
ID ABK40235 standard; cDNA; 1551 BP.
XX
AC ABK40235;
XX
DT 21-MAY-2002 (first entry)
XX
DE DNA encoding human glucocerebrosidase.
XX
KW Human; glucocerebrosidase; GCB; antidiabetic; cytostatic; cancer;
KW antiinflammatory; dermatological; immunosuppressive; antirheumatic;
KW antiarthritic; antithyroid; detergent; textile; food product;
KW agrochemical; feed product; viral pathogen; parasite; autoantigen;
KW autoimmune disease; insulin-dependent diabetes mellitus; gene therapy;
KW autoimmune thyroid disorder; inflammatory condition; sperm antigen;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200202597-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-DK00459.
XX
PR 30-JUN-2000; 2000DK-0001027.
PR 14-JUL-2000; 2000DK-0001092.
PR 29-DEC-2000; 2000WO-DK00743.
PR 09-FEB-2001; 2001WO-DK00090.
XX
PA (MAXY-) MAXYGEN APS.
PA (MAXY-) MAXYGEN HOLDINGS LTD.
XX
PI Okkels JS, Jensen AD, Van Den Hazel B;
XX
DR WPI; 2002-195745/25.
DR P-PSDB; AAU86020.
XX
PT Novel glycosylated polypeptides having improved properties as compared
PT to polypeptide of interest e.g. improved shelf-life, stability, and
PT reduced renal clearance, comprise an additional glycosylation site -
XX
PS Example 1; Page 92; 136pp; English.
XX
CC The invention relates to novel glycosylated polypeptides (I) into which
CC an additional glycosylation site (GS) has been introduced. (I) comprises
CC the primary structure: NH2-X-Pp-COOH or NH2-Px-X-Py-COOH where:
CC X = a peptide addition comprising or contributing to a GS; Pp = a
CC polypeptide; Px = an N-terminal part of Pp; and Py = a C-terminal of Pp.
CC Also described is a method of improving one or more selected properties
CC of Pp, by preparing a nucleotide sequence encoding a polypeptide with the
CC primary structure NH2-X-Pp-COOH, where X is a peptide addition comprising
CC or contributing to a GS that is capable of conferring the selected
CC improved properties to Pp, expressing the nucleotide sequence in a
CC suitable host cell, optionally conjugating the polypeptide to a
CC second non-peptide moiety and recovering the polypeptide. The
CC glycosylated polypeptides have improved properties compared to the
CC plasma protein, erythrocyte or thrombocyte protein, cytokine, growth
CC factor, profibrinolytic protein, protease inhibitor, antigen, enzyme,
CC ligand, receptor, or a hormone, or is a microbial enzyme chosen from
CC protease, amylase, amyloglucosidase, pectinase, lipase and cutinase. Pp
CC can be a therapeutic polypeptide useful in human or veterinary therapy, a

CC diagnostic polypeptide used in diagnostics, an industrial polypeptide
CC useful in the manufacture of products such as detergents, household
CC articles, textile, food products, agrochemicals, feed products,
CC polypeptides or proteins associated with viral pathogens, antigens from
CC parasites, autoantigens associated with autoimmune diseases such as
CC insulin-dependent diabetes mellitus; autoimmune thyroid disorders;
CC inflammatory conditions; sperm antigens; and bacterial and cancer
CC antigens. When the polypeptide is an in vivo glycosylated
CC polypeptide which does not comprise any other type of non-peptide
CC moiety, a nucleotide sequence encoding the polypeptide can be used in
CC gene therapy. When the polypeptide Pp is an antigen, the polypeptide
CC can be provided in the form of a vaccine. ABK40235-ABK40253 represent
CC human glucocerebrosidase coding sequences, PCR primers and related
CC coding sequences of the invention.

xx
SQ Sequence 1551 BP; 345 A; 470 C; 389 G; 347 T; 0 other;

Query Match 67.7%; Score 1543; DB 24; Length 1551;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTGGGCATCAGGTGCC 225
Db 1 ATGGCTGGCAGCCTCACAGGATTGCTTCTACTTTCAGGCAGTGTGGGCATCAGGTGCC 60

QY 226 CGCCCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 285
Db 61 CGCCCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 120

QY 286 TACTGTGACTCCTTTGACCCCCCGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAG 345
Db 121 TACTGTGACTCCTTTGACCCCCCGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAG 180

QY 346 AGTACACGAGTGGCGGACGGATGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACG 405
Db 181 AGTACACGAGTGGCGGACGGATGAGTATGGGGCCCATCCAGGCTAATCACACG 240

QY 406 GGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAAAGTTCCAGAAAGTGAAGGATTT 465
Db 241 GGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAAAGTTCCAGAAAGTGAAGGATTT 300

QY 466 GGAGGGGCATGACAGATGCTGTGCTCTCAACATCCTTGGCCTGTACCCCTGCCCAA 525
Db 301 GGAGGGGCATGACAGATGCTGTGCTCTCAACATCCTTGGCCTGTACCCCTGCCCAA 360

QY 526 AATTGCTACTTAAATCGTACTTCTCTGAAGAAAGGAATCGATATAACATCATCCGGTA 585
Db 361 AATTGCTACTTAAATCGTACTTCTCTGAAGAAAGGAATCGATATAACATCATCCGGTA 420

QY 586 CCCATGGCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGTATGAT 645
Db 421 CCCATGGCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGTATGAT 480

QY 646 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
Db 481 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540

QY 706 CACCGAGCCTGCAGTTGGCCCCAGCGTCCCGTTTCACTCCTTCCAGCCCCCTGGACATCA 765
Db 541 CACCGAGCCTGCAGTTGGCCCCAGCGTCCCGTTTCACTCCTTCCAGCCCCCTGGACATCA 600

QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 825
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 660

QY 826 GGAGACATCTACCACAGACCTGGGCCAGATACCTTGTGAAGTTCCTGGATGCCATGCT 885
Db 661 GGAGACATCTACCACAGACCTGGGCCAGATACCTTGTGAAGTTCCTGGATGCCATGCT 720

QY 886 GAGCACAACTTACAGTTCTGGCAGTACAGCTGAAATGAGCCTTCTGTGGCTGTTG 945
Db 721 GAGCACAACTTACAGTTCTGGCAGTACAGCTGAAATGAGCCTTCTGTGGCTGTTG 780

QY 946 AGTGGATACCCCTTCCAGTGCCTGGGCTTACCCTTGAACATCAGCAGACTTCATTGCC 1005
Db 781 AGTGGATACCCCTTCCAGTGCCTGGGCTTACCCTTGAACATCAGCAGACTTCATTGCC 840

QY 1006 CGTGACCTAGTCTTACCCTCGCCAAACAGTACTCACCACAATGTCGCTACTCATGCTG 1065
Db 841 CGTGACCTAGTCTTACCCTCGCCAAACAGTACTCACCACAATGTCGCTACTCATGCTG 900

QY 1066 GATGACCAACGCTTGTGCTGCTGCCCTTGGGCAAGGTGGTACTGACAGACCCAGAAGCA 1125
Db 901 GATGACCAACGCTTGTGCTGCTGCCCTTGGGCAAGGTGGTACTGACAGACCCAGAAGCA 960

QY 1126 GCTAAATATGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1185
Db 961 GCTAAATATGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1186 GCCACCTTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTCTTTCCTTCAGAGGCC 1245
Db 1021 GCCACCTTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTCTTTCCTTCAGAGGCC 1080

QY 1246 TGTGTGGGCTCCAAGTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATG 1305
Db 1081 TGTGTGGGCTCCAAGTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATG 1140

QY 1306 CAGTACAGCCACAGCATCATCAGAACCTCTGTACCATGTGGTGGCTGGACCGACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGAACCTCTGTACCATGTGGTGGCTGGACCGACTGG 1200

QY 1366 AACCTTGGCTGAAACCCGAGGAGGACCCCAATTTGGTGGTAACTTTGTGACAGTCCC 1425
Db 1201 AACCTTGGCTGAAACCCGAGGAGGACCCCAATTTGGTGGTAACTTTGTGACAGTCCC 1260

QY 1426 ATCATTTGTAGACATCACCAGGACACACGTTTACAAACAGCCCATGTTTACCACCTTGGC 1485
Db 1261 ATCATTTGTAGACATCACCAGGACACACGTTTACAAACAGCCCATGTTTACCACCTTGGC 1320

QY 1486 CACTTCAGCAAGTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAAGAAG 1545
Db 1321 CATTTCAGCAAGTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAAGAAG 1380

QY 1546 AACGACCTGGACGAGTGGCAGTGCCTCTTACCATCAAGGATCCTGCTGCTGGGCTTCTCTGGAG 1605
Db 1381 AACGACCTGGACGAGTGGCAGTGCCTCTTACCATCAAGGATCCTGCTGCTGGGCTTCTCTGGAG 1440

QY 1606 AACCGCTCCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGCTGGGCTTCTCTGGAG 1665
Db 1441 AACCGCTCCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGCTGGGCTTCTCTGGAG 1500

QY 1666 ACAATCTCACCTGCTACTTCCATTTCACACCTTACCTGTGGCGTCCGACAGTGA 1716
Db 1501 ACAATCTCACCTGCTACTTCCATTTCACACCTTACCTGTGGCGTCCGACAGTGA 1551

RESULT 12
AAH99924
ID AAH99924 standard; cdna; 1592 BP.
XX
AC AAH99924;
XX
DT 13-FEB-2002 (first entry)
XX
DE Nucleotide sequence of human glucocerebrosidase.
XX
KW Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
KW Farber disease; Tay-sachs disease; Niemann-pick disease;
KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
KW Gaucher disease; glycosylation; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
13..1563
/*tag= a

FT XX /product= "Human glucocerebrosidase"
PN WO200177307-A2.
XX 18-OCT-2001.
PD 06-APR-2001; 2001WO-US11144.
XX 06-APR-2000; 2000US-195598P.
PF (CYTO-) CYTOCLONAL PHARM INC.
XX Berent SL;
PI WPI; 2002-041292/05.
XX P-PSDB; AAG78410.
DR Novel expression system useful for producing glucocerebrosidase for
XX treating Gaucher's disease, comprises an insect cell transformed with a
PT vector encoding glucocerebrosidase that synthesizes glucocerebrosidase
PT
XX
PS Claim 2; Page 62-64; 74pp; English.
XX This invention relates to an expression system comprising an insect
CC cell transformed with a vector encoding glucocerebrosidase (GC) that
CC synthesizes clinically effective GC. Recombinant GC is useful for
CC treating individuals with deficiencies in GC, by introducing the
CC clinically effective form of recombinant GC from insect cells into
CC individuals. The expression system is stably transformed and
CC provides a consistently higher level of expression of GC than in
CC baculovirus or mammalian cell expression systems, and proper
CC glycosylation modifications for GC, requiring no enzymatic
CC carbohydrate remodeling to be clinically effective. The expression
CC system is more effective, economical and simple for the manufacture of
CC recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack
CC of lysosomal enzymes contributes to some of the many genetically
CC inherited lysosomal storage diseases, such as Fabry disease, Farber
CC disease, Tay-Sachs disease, Niemann-Pick disease, Hunter syndrome,
CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease.
CC This nucleotide sequence represents that of human
CC glucocerebrosidase. See also AAH99925 for the nucleotide sequence of
CC human glucocerebrosidase, noting the differences at residues
CC 1552 to 1554. This sequence encodes arginine at that position,
XX AAH99925 encodes histidine.
SQ Sequence 1592 BP; 359 A; 481 C; 397 G; 355 T; 0 other;
Query Match 67.1%; Score 1529.6; DB 24; Length 1592;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1541; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 157 GTAAGCATCATGGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCA 216
DB 4 GAATTCAAAATGGCTGGCTCGTTAACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCA 63
QY 217 TCAGGTGCCCGCCCTGTCATCCCTTAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTGC 276
DB 64 TCCGGAGCTAGACCTTGCATCCCTTAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTGC 123
QY 277 AATGCCACATACCTGTGACTCCTTTGACCCCCCGACCTTTCCTGCCCTTGGTACCTTCAGC 336
DB 124 AATGCCACATACCTGTGACTCCTTTGACCCCCCGACCTTTCCTGCCCTTGGTACCTTCAGC 183
QY 337 CGTATGAGAGTACACGCGAGTGGCGACGGATGAGCTGAGTATGGGGCCCATCCAGGCT 396
DB 184 CGCTATGAGAGTACACGCGAGTGGCGACGGATGAGCTGAGTATGGGGCCCATCCAGGCT 243
QY 397 AATCACACGGGACAGGCTGCTACTGACCCCTGACGCCAGAACAGAGTTCAGAAAGTG 456
DB 244 AATCACACGGGACAGGCTGCTACTGACCCCTGACGCCAGAACAGAGTTCAGAAAGTG 303
QY 457 AAGGGATTGGAGGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCCTGTACACC 516

DB 304 AAGGGATTGGAGGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCCTGTACACC 363
QY 517 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAACATC 576
DB 364 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAACATC 423
QY 577 ATCCGGGTACCCCATGGCCAGCTGTGACTTCTCTCCATCCGCACTACACCTATGCAGACACC 636
DB 424 ATCCGGGTACCCCATGGCCAGCTGTGACTTCTCTCCATCCGCACTACACCTATGCAGACACC 483
QY 637 CCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 696
DB 484 CCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 543
QY 697 CCCCTGATTCACCGAGCCCTGCACTTGGCCAGCGTCCCGTTTCACTCCTTGCCAGCCCC 756
DB 544 CCCCTGATTCACCGAGCCCTGCACTTGGCCAGCGTCCCGTTTCACTCCTTGCCAGCCCC 603
QY 757 TGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAG 816
DB 604 TGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAG 663
QY 817 GGACAGCCGGGAGACATCTACCAACCAGACCTGGCCAGATACTTTGTGAAGTTCTCTGGAT 876
DB 664 GGACAGCCGGGAGACATCTACCAACCAGACCTGGCCAGATACTTTGTGAAGTTCTCTGGAT 723
QY 877 GCCTATGCTGAGCACAACTTACAGTTCTTGGGCAGTGACAGCTGAAATGAGCCTTCTGCT 936
DB 724 GCCTATGCTGAGCACAACTTACAGTTCTTGGGCAGTGACAGCTGAAATGAGCCTTCTGCT 783
QY 937 GGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCTTGAACATCAGCGAGAC 996
DB 784 GGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCTTGAACATCAGCGAGAC 843
QY 997 TTCATTGCCCGTGAACCTTAGGTCTTACCTCGCCCAACAGTACTCACCAATGTCCGCCCTA 1056
DB 844 TTCATTGCCCGTGAACCTTAGGTCTTACCTCGCCCAACAGTACTCACCAATGTCCGCCCTA 903
QY 1057 CTCATGCTGGATGACCAACGCTTGTGCTGCCCGCCACATGGGCAAGGTGCTGACAGAC 1116
DB 904 CTCATGCTGGATGACCAACGCTTGTGCTGCCCGCCACATGGGCAAGGTGCTGACAGAC 963
QY 1117 CCAGAACGAGCTAAATATGTTTCATGGCATGCTGTACATGGTACCTGGACTTCTGCTGCT 1176
DB 964 CCAGAACGAGCTAAATATGTTTCATGGCATGCTGTACATGGTACCTGGACTTCTGCTGCT 1023
QY 1177 CCAGCAAAAGCCACCTTAGGGGAGACACACCGCTGTCTCCCAACACATGCTCTTTGCC 1236
DB 1024 CCAGCAAAAGCCACCTTAGGGGAGACACACCGCTGTCTCCCAACACATGCTCTTTGCC 1083
QY 1237 TCAGAGCCCTGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGAT 1296
DB 1084 TCAGAGCCCTGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGAT 1143
QY 1297 CGAGGATGAGTACAGCCACAGCATATCACGAACCTCCTGTACCATGTGGTGGCTGG 1356
DB 1144 CGAGGATGAGTACAGCCACAGCATATCACGAACCTCCTGTACCATGTGGTGGCTGG 1203
QY 1357 ACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCCAATGGGTGCGTAACCTTGTG 1416
DB 1204 ACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCCAATGGGTGCGTAACCTTGTG 1263
QY 1417 GACAGTCCCATCATTTAGACATCAACAAAGGACACCTTTTACAAACAGCCCATGTTCTAC 1476
DB 1264 GACAGTCCCATCATTTAGACATCAACAAAGGACACCTTTTACAAACAGCCCATGTTCTAC 1323
QY 1477 CACCTTGGCCACTTCAGCAAGTTCAATCCTGAGGGTCCAGAGAGTGGGGTGGTGGCC 1536
DB 1324 CACCTTGGCCACTTCAGCAAGTTCAATCCTGAGGGTCCAGAGAGTGGGGTGGTGGCC 1383
QY 1537 AGTCAGAAAGAACGACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTGTG 1596

Db 1384 AGTCAGAAGAACGACCTGGACGACGAGTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTG 1443
QY 1597 GTCGTGCTAAACCGCTCTCTAAGGATGTGCTCTTACCATCAAGATCCTGCTGTGGGC 1656
Db 1444 GTCGTGCTAAACCGCTCTCTAAGGATGTGCTCTTACCATCAAGATCCTGCTGTGGGC 1503
QY 1657 TTCCTGGAGACAATCTCACCTGGCTACTCCATTCACACCTACCTGCGGCGTCCGAGTGA 1716
Db 1504 TTCCTGGAGACAATCTCACCTGGCTACTCCATTCACACCTACCTGCGGCGTCCGAGTGA 1563

RESULT 13

AAQ39286
ID AAQ39286 standard; DNA; 7620 BP.

XX AC AAQ39286;

XX DT 20-JUL-1993 (first entry)

XX DE Glucocerebrosidase gene.

XX KW Glucocerebrosidase; peripheral blood leukocyte; lysosomal degradation;
KW glycolipid; Gaucher disease; glucosylceramide; glucocerebrosidase; RFLP;
KW restriction fragment length polymorphism; mutation; pseudogene; 1226G;
KW Jewish; 1448C; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT exon 356..611

FT exon /*tag= a

FT exon /number= Exon 1

FT intron 612..979

FT intron /*tag= b

FT intron /number= Intron 1

FT exon 980..1067

FT exon /*tag= c

FT intron /number= Exon 2

FT intron 1068..1619

FT intron /*tag= d

FT intron /number= Intron 2

FT exon 1620..1811

FT exon /*tag= e

FT intron /number= Exon 3

FT intron 1812..1934

FT intron /*tag= f

FT exon /number= Intron 3

FT exon 1935..2081

FT exon /*tag= g

FT intron /number= Exon 4

FT intron 2082..3046

FT intron /*tag= h

FT exon /number= Intron 4

FT exon 3047..3180

FT exon /*tag= i

FT intron /number= Exon 5

FT intron 3181..3390

FT intron /*tag= j

FT exon /number= Intron 5

FT exon 3391..3563

FT exon /*tag= k

FT intron /number= Exon 6

FT intron 3564..4116

FT intron /*tag= l

FT exon /number= Intron 6

FT exon 4117..4354

FT exon /*tag= m

FT intron /number= Exon 7

FT intron 4355..5227

FT intron /*tag= n

FT exon /number= Intron 7

FT exon 5228..5452

FT exon /*tag= o

FT intron /number= Exon 8
FT intron 5453..5852
FT exon /*tag= p
FT exon /number= Intron 8
FT exon 5853..6016
FT exon /*tag= q
FT intron /number= Exon 9
FT intron 6017..6385
FT exon /*tag= r
FT exon /number= Intron 9
FT exon 6386..6502
FT exon /*tag= s
FT intron /number= Exon 10
FT intron 6503..6596
FT exon /*tag= t
FT exon /number= Intron 10
FT exon 6597..7245
FT exon /*tag= u
FT exon /number= Exon 11

XX WO9306244-A.

XX PD 01-APR-1993.

XX PF 16-SEP-1992; 92WO-US07840.

XX PR 27-SEP-1991; 91US-0767135.

XX PA (SCRI) SCRIPPS RES INST.

XX PI Beutler E, Sorge JA;

XX DR WPI; 1993-117560/14.

XX PT Screening method for new Gaucher disease mutation - comprises
XX PT amplifying DNA by PCR and assaying the prod. with probe to
XX PT guanine insertion adjacent position 57 of gluco-cerebrosidase gene-
XX PT exon 2

XX PS Disclosure; Page 55-60; 74pp; English.

XX CC This sequence represents the glucocerebrosidase gene which was
XX CC isolated from peripheral blood leukocytes. Glucocerebrosidase is an
XX CC enzyme which is required for the lysosomal degradation of glycolipids.
XX CC A deficiency of this enzyme leads to Gaucher disease, as in the absence
XX CC of glucocerebrosidase, the extremely insoluble glucosylceramide
XX CC (glucocerebroside) accumulates. The gene for glucocerebrosidase is
XX CC located on chromosome 1, q21. A number of different mutations cause
XX CC Gaucher disease and analysis of these mutations has been complicated
XX CC by the presence of a pseudogene 16 kb downstream from the glucocere-
XX CC brosidase gene. The pseudogene is approx. 95% homologous to the
XX CC functional gene and can be transcribed but not translated. About
XX CC 75% of Jewish patients with clinically significant Gaucher disease
XX CC contain a characteristic A>G mutation at cDNA nucleotide 1226 (1226G
XX CC mutation) which corresponds to amino acid 370 of the mature protein.
XX CC The corresponding position of the mutation in the glucocerebrosidase
XX CC gene is in exon 9, position 2. The same mutation is found in approx.
XX CC 25% of non-Jewish disease causing alleles. This mutation causes a
XX CC characteristic restriction fragment length polymorphism (RFLP). A
XX CC second, less common mutation is at cDNA position 1448 and is a T>C
XX CC mutation (1448C mutation). The corresponding position of this
XX CC mutation is exon 10, position 60. The 1448C mutation only accounts
XX CC for 2% of Jewish Gaucher disease producing alleles and for about 40%
XX CC of the alleles in non-Jewish patients.

XX SQ Sequence 7620 BP; 1665 A; 2091 C; 1925 G; 1939 T; 0 other;

Query Match 29.3%; Score 667; DB 14; Length 7620;
Best Local Similarity 89.1%; Pred. No. 1.1e-175;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;
QY 1489 TTCAGCAAGTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGCTGCCAGTCAAGAAC 1548
|||||

[illegible]

RESULT 14

ABL67250

ID ABL67250 standard; DNA; 5769 BP.

XX
XX
XXXXX

AC ABL67250;

XX XX

DT 15-MAY-2002 (first entry)

XXI

Thyroid cancer related genes

XX

KW Human; cancer; colon; brea

| | | |
|----|------------------------------|--|
| XX | Homo sapiens. | |
| OS | WO200194629-A2. | |
| XX | 13-DEC-2001. | |
| PN | 30-MAY-2001; 2001WO-US10838. | |
| XX | 05-JUN-2000; 2000US-209473P. | |
| PF | 05-JUN-2000; 2000US-209531P. | |
| XX | 18-SEP-2000; 2000US-233133P. | |
| PD | 18-SEP-2000; 2000US-233617P. | |
| XX | 20-SEP-2000; 2000US-234009P. | |
| XX | 20-SEP-2000; 2000US-234034P. | |
| XX | 20-SEP-2000; 2000US-234052P. | |
| XX | 22-SEP-2000; 2000US-234509P. | |
| XX | 22-SEP-2000; 2000US-234567P. | |
| XX | 25-SEP-2000; 2000US-234923P. | |
| XX | 25-SEP-2000; 2000US-234924P. | |
| XX | 25-SEP-2000; 2000US-235077P. | |
| XX | 25-SEP-2000; 2000US-235082P. | |
| XX | 25-SEP-2000; 2000US-235134P. | |
| XX | 25-SEP-2000; 2000US-235280P. | |
| XX | 26-SEP-2000; 2000US-235637P. | |
| XX | 26-SEP-2000; 2000US-235638P. | |
| XX | 27-SEP-2000; 2000US-235711P. | |
| XX | 27-SEP-2000; 2000US-235720P. | |
| XX | 27-SEP-2000; 2000US-235840P. | |
| XX | 27-SEP-2000; 2000US-235863P. | |
| XX | 28-SEP-2000; 2000US-236028P. | |
| XX | 28-SEP-2000; 2000US-236032P. | |
| XX | 28-SEP-2000; 2000US-236033P. | |
| XX | 28-SEP-2000; 2000US-236034P. | |
| XX | 28-SEP-2000; 2000US-236109P. | |
| XX | 28-SEP-2000; 2000US-236111P. | |
| XX | 29-SEP-2000; 2000US-236842P. | |
| XX | 29-SEP-2000; 2000US-236891P. | |
| XX | 02-OCT-2000; 2000US-237172P. | |
| XX | 02-OCT-2000; 2000US-237173P. | |
| XX | 02-OCT-2000; 2000US-237178P. | |
| XX | 02-OCT-2000; 2000US-237294P. | |
| XX | 02-OCT-2000; 2000US-237295P. | |
| XX | 02-OCT-2000; 2000US-237316P. | |
| XX | 03-OCT-2000; 2000US-237425P. | |
| XX | 03-OCT-2000; 2000US-237598P. | |
| XX | 03-OCT-2000; 2000US-237604P. | |
| XX | 03-OCT-2000; 2000US-237606P. | |
| XX | 03-OCT-2000; 2000US-237608P. | |
| XX | 01-NOV-2000; 2000US-244867P. | |
| XX | 01-NOV-2000; 2000US-245084P. | |

(AVAL-) AVALON PHARM.

Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
Soppet DR, Weaver Z;

WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene

Claim 1; SEQ ID 5587; 44pp: English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening

CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 5769 BP; 1285 A; 1566 C; 1476 G; 1436 T; 6 other;

Query Match 28.1%; Score 639.4; DB 24; Length 5769;
Best Local Similarity 87.6%; Pred. No. 5.2e-168;
Matches 758; Conservative 0; Mismatches 12; Indels 95; Gaps 2;

QY 1489 TTCAGCAAGTTTCATCTCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAGAGAAGAC 1548
DB 4792 TTCAGCAAGTTTCATCTCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAGAGAAGAC 4851
QY 1549 GACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTGTGCTGCTAAAC 1608
DB 4852 GACCCGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTGTGCTGCTAAAC 4911
QY 1609 C----- 1609
DB 4912 CGGTGAGGGCAATGCTGAGGTCTGGGAAGTGGGCTGAAGACAGCGTGGGGCCTTGGCA 4971
QY 1610 -----GCTCCTCTAAGGATGCTCTTAC 1634
DB 4972 GGATCACACTCTCAGCTTCTCCTCCCTGCTCCTAGTCTCTTAAGGATGCTCTTAC 5031
QY 1635 CATCAAGGATCCTGCTGCTGGGCTTCTCTGGAGACAATCTCACCTGGCTACTCCATTACAC 1694
DB 5032 CATCAAGGATCCTGCTGCTGGGCTTCTCTGGAGACAATCTCACCTGGCTACTCCATTACAC 5091
QY 1695 CTACCTGTGGCGTCCGAGTGGAGGAGTCAAGGAGGCACTGGGCTCAGCCTGG 1754
DB 5092 CTACCTGTGGCGTCCGAGTGGAGGAGTCAAGGAGGCACTGGGCTCAGCCTGG 5151
QY 1755 GCATTAAAGGACAGAGTCACTCACACGCTGCTGTGACTAAAGAGGACAGCAGGGC 1814
DB 5152 GCATTAAAGGACAGAGTCACTCACACGCTGCTGTGACTAAAGAGGACAGCAGGGC 5211
QY 1815 CAGTGTGAGCTTACAGGACGTAAGCCAGGGCAATGTTGGTGGTCACTCTTCCC 1874
DB 5212 CAGTGTGAGCTTACAGGACGTAAGCCAGGGCAATGTTGGTGGTCACTCTTCCC 5271
QY 1875 TCTAGTGGTGGCAGGGCTGGAGGCGCTTGTGCTGCTGCTTGTGCTTGTGCTGCTGCTG 1934
DB 5272 TCTAGTGGTGGCAGGGCTGGAGGCGCTTGTGCTGCTGCTTGTGCTTGTGCTGCTGCTG 5331
QY 1935 CCAGCCCCCATGCTTATGTGAACATGCGCTGCTGCTGCTTGTGCTTGTGCTGCTGCTG 1994
DB 5332 CCAGCCCCCATGCTTATGTGAACATGCGCTGCTGCTGCTTGTGCTTGTGCTGCTGCTG 5390
QY 1995 GGTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACACAGATCAGGGCTGAGGTAAG 2054
DB 5391 GGTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACACAGATCAGGGCTGAGGTAAG 5450
QY 2055 GAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTGCTTCTTCTGGAGA 2114
DB 5451 GAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTGCTTCTTCTGGAGA 5510
QY 2115 TGCAGAACTGGGCCCCGTGGAGCAGCAGTGTCAAGTCAAGGGCGGAAGCCTTAAAGCAGCA 2174
DB 5511 TNNNNNCTGGGCCCCGTGGAGCAGCAGTGTCAAGTCAAGGGCGGAAGCCTTAAAGCAGCA 5570
QY 2175 GCGGGTGTGCCAGGACCCAGATGATTCCTATGGCACCCAGCCAGGAAATGGCAGCTC 2234
DB 5571 GCGGGTGTGCCAGGACCCAGATGATTCCTATGGCACCCAGCCAGGAAATGGCAGCTC 5630
QY 2235 TTAAAGGAGAAAAATGTTTGAGCCCA 2259

DB 5631 TTAAAGGAGAAAAATGTTTGAGCCCA 5655

RESULT 15

AAC00312

ID AAC00312 standard; cDNA; 417 BP.

XX AAC00312;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 310.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
PN 06-SEP-2000.
PD 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
PR (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX P-PSDB; AAG00306.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 1; SEQ ID 310; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from
mRNAs encoding secreted proteins. An ORF has been identified within the
sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
derived from 30 different tissues. EST sequences usually correspond
mainly to the 3' untranslated region (UTR) of the mRNA because they are
often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
well suited for isolating cDNA sequences derived from the 5' ends of
mRNAs and even in those cases where longer cDNA sequences have been
obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
mRNAs with intact 5' ends and can therefore be used to obtain full length
cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
gene therapy and chromosome mapping procedures. They are used to obtain
upstream regulatory sequences and to design expression and secretion
vectors.

Sequence 417 BP; 81 A; 114 C; 109 G; 112 T; 1 other;

Query Match 15.9%; Score 363; DB 21; Length 417;
Best Local Similarity 99.7%; Pred. No. 3.2e-91;
Matches 374; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AGCTAAGGCAGGTACCTGCATCTCTGTTTGTAGTGGATCCTCTATCCTTCAGAGAC 60

DB 44 AGCTAAGGCAGGTACCTGCATCTCTGTTTGTAGTGGATCCTCTATCCTTCAGAGAC 103

QY 61 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGT 120

DB 104 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGT 163

QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGGTGGCAGCTC 180

DB 164 CCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGGTGGCAGCTC 223

| | | | |
|----|-----|---|-----|
| Qy | 181 | ACAGGATTGCTTCTACTTCAGGCAGTGTCGTGGGCATCAGGTGCCCGCCCCCTGCATCCCT | 240 |
| Db | 224 | ACAGGATTGCTTCTACTTCAGGCAGTGTCGTGGGCATCAGGTGCCCGCCCCCTGCATCCCT | 283 |
| Qy | 241 | AAAAGCTTCGGCTACAGCTCGGTGGTGTTGTCTGCAATGCCACATACTGTGACTCCTTT | 300 |
| Db | 284 | AAAAGCTTCGGCTACAGCTCGGTGGTGTTGTCTGCAATGCCACATACTGTGACTCCTTT | 343 |
| Qy | 301 | GACCCCCCGACCTTTCCTGCCCTTGGTACCCTCAGCCGCTATGAGAGTACACGCAGTGG | 360 |
| Db | 344 | GACCCCCCGACCTTTCCTGCCCTTGGTACCCTCAGCCGCTATGAGAGTACACGCAGT-GG | 402 |
| Qy | 361 | CGACGGATGGAGCTG | 375 |
| Db | 403 | CGACGGATGGAGCTG | 417 |

Search completed: February 22, 2003, 10:41:50
Job time : 535 secs

10

10

10

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 22, 2003, 10:29:56 ; Search time 2925 Seconds
(without alignments)
12618.637 Million cell updates/sec

Title: US-10-024-197-24
Perfect score: 2279
Sequence: 1 agtaagcaggtactctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 1361.8 | 59.8 | 1931 | 11 | BC030240 |
| 2 | 921.8 | 40.4 | 953 | 9 | AL541128 |
| c 3 | 901.6 | 39.6 | 915 | 9 | AL571401 |
| c 4 | 896.2 | 39.3 | 927 | 9 | AL568749 |
| 5 | 879.8 | 38.6 | 940 | 9 | AL524792 |
| 6 | 864.6 | 37.9 | 920 | 14 | BQ676149 |

| | | | | | |
|------|-------|------|------|----|----------|
| 7 | 858.4 | 37.7 | 916 | 12 | BG744064 |
| c 8 | 856.8 | 37.6 | 957 | 12 | BG744733 |
| c 9 | 850.2 | 37.3 | 909 | 9 | AL562392 |
| 10 | 841 | 36.9 | 931 | 14 | BQ680411 |
| 11 | 835 | 36.6 | 838 | 14 | BQ681028 |
| 12 | 833.4 | 36.6 | 916 | 14 | BQ677036 |
| 13 | 830.8 | 36.5 | 1017 | 13 | BM477383 |
| 14 | 829 | 36.4 | 896 | 14 | BQ897517 |
| 15 | 826 | 36.2 | 1008 | 14 | BQ69685 |
| 16 | 824 | 36.2 | 938 | 14 | BQ683279 |
| 17 | 812.6 | 35.7 | 872 | 14 | BQ678499 |
| 18 | 811.8 | 35.6 | 896 | 14 | BQ678522 |
| 19 | 811.2 | 35.6 | 863 | 12 | BQ281198 |
| 20 | 811 | 35.6 | 896 | 14 | BQ683460 |
| c 21 | 805.4 | 35.3 | 849 | 9 | AL562243 |
| 22 | 801 | 35.1 | 954 | 9 | AL523699 |
| 23 | 794.8 | 34.9 | 945 | 12 | BG674486 |
| 24 | 791.4 | 34.7 | 923 | 14 | BQ680695 |
| 25 | 779.8 | 34.2 | 837 | 13 | BI917380 |
| 26 | 766.6 | 33.6 | 861 | 12 | BG678590 |
| 27 | 763.2 | 33.5 | 1043 | 14 | BQ071624 |
| 28 | 754.8 | 33.1 | 793 | 13 | BI910946 |
| c 29 | 748.6 | 32.8 | 775 | 14 | BQ008716 |
| c 30 | 748.4 | 32.8 | 774 | 14 | BQ574459 |
| 31 | 738 | 32.4 | 787 | 9 | AU122294 |
| 32 | 735 | 32.3 | 862 | 13 | BI756617 |
| 33 | 735 | 32.3 | 913 | 12 | BG765009 |
| 34 | 734.8 | 32.2 | 855 | 13 | BI769458 |
| 35 | 731.8 | 32.1 | 1070 | 13 | BM544110 |
| c 36 | 729.2 | 32.0 | 763 | 14 | BQ000311 |
| c 37 | 725.8 | 31.8 | 795 | 9 | AL572439 |
| 38 | 722.6 | 31.7 | 873 | 14 | BQ425733 |
| 39 | 722.4 | 31.7 | 725 | 14 | BM772392 |
| 40 | 719.2 | 31.6 | 859 | 12 | BG764893 |
| c 41 | 715.2 | 31.4 | 756 | 10 | BE613411 |
| 42 | 713.4 | 31.3 | 773 | 9 | AU138949 |
| 43 | 713 | 31.3 | 885 | 12 | BG749231 |
| 44 | 710.6 | 31.2 | 814 | 9 | AU140959 |
| 45 | 707.4 | 31.0 | 885 | 12 | BE867971 |

ALIGNMENTS

RESULT 1
BC030240
LOCUS BC030240
DEFINITION Homo sapiens, Similar to glucosylase, beta; acid (includes glucosylceramidase), clone IMAGE:5114119, mRNA.
ACCESSION BC030240
VERSION BC030240.1
KEYWORDS GI:20988849
SOURCE HTC.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 1931)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Laric, P., Legaspi, R., Maduro, Q.L.,
Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C.,
Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 62 Row: c Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4503934
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..1931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5114119"
/tissue_type="Cervix, carcinoma"
/clone_lib="NIH_MGC_12"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 466 a 547 c 515 g 403 t

ORIGIN

Query Match 59.8%; Score 1361.8; DB 11; Length 1931;
Best Local Similarity 81.0%; Pred. No. 5.7e-259;
Matches 1813; Conservative 0; Mismatches 87; Indels 339; Gaps 6;

QY 37 GTGGATCCTCTATCCTTCAGAGACTCTGGAAACCCCTGTGGTCTCTCTTCTCATCTAATGAC 96
|||||
DB 32 GTGGATCCTCTATCCTTCAGAGACTCTGGAAACCCCTGTGGTCTCTCTCTCATCTAGTGAC 91
QY 97 CCTGAGGGATGGAGTTTCAAGTCTCTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGG 156
|||||
DB 92 CCTGAGGGATGGAGTTTCAAGTCTCTCCAGAGAGGAATGTCCCAAGCCTTCGGGTAGG 151
QY 157 GTAAGCATATGGCTGGCAGCCTCACAGGATGCTTCTACTCAGGCAGTGTCTGGGCA 216
|||||
DB 152 GTAAGCATATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTCTGGGCA 211
QY 217 TCAGGTGCCGCCCTGCATCCCTTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGC 276
|||||
DB 212 TCAGATG----- 218
QY 277 AATGCCACATATGTGACTCCTTTGACCCCCCGACCTTTCTCTGCCCTTGGTACCTTCAGC 336
|||||
DB 219 -----AGTGCCTTCAGC 230
QY 337 CGCTATGAGATACAGCAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCT 396
|||||
DB 231 CGCTACAGAGAGAGCAGTGGGATGGAGTGGAGCTGAGTA-CAGGACCATACAGGCT 289
QY 397 AATCACACGGSCACAGGCTGTCTACTGACCTGCAGCCAGAGAAGTTCCAGAAAAGTG 456
|||||
DB 290 AATTGCACCGGCACA----- 304
QY 457 AAGGGATTGGAGGGGCCATGACAGATGCTGTCTCTCAACATCCTTGGCCCTGTACCC 516
|||||
DB 305 -----GTGAAATAAGATTTCGCCGCCCTATCATCTGCTAACTGCTACGGACT 350
QY 517 CCTGCCCCAAATTTGCTACTTAATCGTACTTCTCTGAAGAGGAATCGGATATAACATC 576
|||||
DB 351 CAGGCTCAGAAAGCCCTGC-----GCTTCAACCCAGGAATCGGATATAACATC 397
QY 577 ATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCCACCTACACCTATGCAGACACC 636
|||||
DB 398 ATCTGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCCACCTACACCTATGCAGACACC 457
QY 637 CCTGATGATTTCCAGTTGCACAACCTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 696
|||||

Db 1373 TGGAGCAGATACCTAAGAGGCACTGGGCTCAGCCTGGGCATTAAGGGACAGAGTCAGC 1432

QY 1777 TCACACGCTGTCTGTGACTAAAGAGGGACAGCAGGGCCAGTGTGAGCTTACAGCGACGT 1836

Db 1433 TCACACGCTGTCTGTGACTAAAGAGGGACAGCAGGGCCAGCTGAGCTTACAGCGACGT 1492

QY 1837 AAGCCCGAGGGCAATGGTGGGTGACTCAGTCTTCCCTCTAGGTGGTGGCAGGGGCTGG 1896

Db 1493 AAGCCCGAGGGCAATGGTGGGTGACTCAGTCTTCCCTCTAGGTGGTGGCAGGGGCTGG 1552

QY 1897 AGGCCCTAGAAAAGATCAGTAAGCCCGAGTGTCCCCCAGCCGCCCTATGCTTATGTGAA 1956

Db 1553 AGGCCCTAGAAAAGATCAGTAAGCCCGAGTGTCCCCCAGCCGCCCTATGCTTATGTGAA 1612

QY 1957 CATGCGCTGTGTGCTGTCTTGGAACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 2016

Db 1613 CATGCGCTGTGTGCTGTCTTGGAACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1672

QY 2017 ACTGTCCTACAAACACAGATCAGGCTGAGGGTAAAGGAAAGAAAGAGACTAGGAAAGC 2076

Db 1673 ACTGTCCTACAAACACAGATCAGGCTGAGGGTAAAGGAAAGAAAGAGACTAGGAAAGC 1732

QY 2077 TGGGCCCAAACTGGAGACTGTTTGTCTTCTTCTGAGATGCAGAACTGGGCCCCGTGGAGC 2136

Db 1733 TGGGCCCAAACTGGAGACTGTTTGTCTTCTTCTGAGATGCAGAACTGGGCCCCGTGGAGC 1792

QY 2137 AGCAGTGTGAGCATCAGGGCGGAAAGCTTAAAGCAGCAGCGGGTGTGCCAGGCAACCCAG 2196

Db 1793 AGCAGTGTGAGCATCAGGGCGGAAAGCTTAAAGCAGCAGCGGGTGTGCCAGGCAACCCAG 1852

QY 2197 ATGATTCTTATGGCACCAGCCAGGAAATGGCAGCTTAAAGGAGAGAAATGTTTGGAGC 2256

Db 1853 ATGATTCTTATGGCACCAGCCAGGAAATGGCAGCTTAAAGGAGAGAAATGTTTGGAGC 1912

QY 2257 CCAAAAAAAAAAAAAAAAAA 2275

Db 1913 CCAAAAAAAAAAAAAAAAAA 1931

RESULT 2

AL541128

LOCUS

DEFINITION AL541128 LTI_FL002_PL1 Homo sapiens cDNA clone EST 16-FEB-2001

ACCESSION AL541128

VERSION AL541128.1 GI:12871896

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. .953

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DE005YL08"

/lab_host="DH10B"

/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville

Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 214 a 271 c 240 g 223 t 5 others

ORIGIN

Query Match 40.4%; Score 921.8; DB 9; Length 953;

Best Local Similarity 99.1%; Pred. No. 5e-172;

Matches 942; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

QY 1 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTGTTAGTGGATCCTCTATCCTTCAGAGAC 60

Db 4 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTGTTAGTGGATCCTCTATCCTTCAGAGAC 63

QY 61 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGGATGGAGTTTCAAGT 120

Db 64 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGGATGGAGTTTCAAGT 123

QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCCTTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180

Db 124 CCTTCCAGAGAGGAATGTCCCAAGCCCTTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 183

QY 181 ACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCATCAGTGCCGCCCTGTCATCCCT 240

Db 184 ACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCATCAGTGCCGCCCTGTCATCCCT 243

QY 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATACTGTGACTCCTTT 300

Db 244 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATACTGTGACTCCTTT 303

QY 301 GACCCCGCAGCTTCTTCCCTGGTGGTGTGTCTGCAATGCCACATACTGTGACTCCTTT 360

Db 304 GACCCCGCAGCTTCTTCCCTGGTGGTGTGTCTGCAATGCCACATACTGTGACTCCTTT 363

QY 361 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGGGCAGGCTGCTA 420

Db 364 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGGGCAGGCTGCTA 423

QY 421 CTGACCTTCAGCCAGCAAGAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACA 480

Db 424 CTGACCTTCAGCCAGCAAGAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACA 483

QY 481 GATGCTGTGCTCTCAACATCCTTCCCTGTGTCACCCCTGCCCCAAAATTTGCTACTTAA 540

Db 484 GATGCTGTGCTCTCAACATCCTTCCCTGTGTCACCCCTGCCCCAAAATTTGCTACTTAA 543

QY 541 TCGTACTTCTCTGAAGAGGAATCGGATATACATCAATCCGGGTACCCATGGCCAGCTGT 600

Db 544 TCGTACTTCTCTGAAGAGGAATCGGATATACATCAATCCGGGTACCCATGGCCAGCTGT 603

QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAAC 660

Db 604 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAAC 663

QY 661 TTCAGCTCCAGAGGAGATACCAAGTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720

Db 664 TTCAGCTCCAGAGGAGATACCAAGTCAAGATACCCCTGATTCACCGAGCCCTGCAG 723

QY 721 TTGGCCAGCGTCCCGTTTCACTCTCTGCGAGCCCTGGACATCACCCACTTGGCTCAAG 780

Db 724 TTGGCCAGCGTCCCGTTTCACTCTCTGCGAGCCCTGGACATCACCCACTTGGCTCAAG 783

QY 781 ACCAATGGAGCGGTGAATGGGAGGGTCACTCAAGGACAGCCCGGAGACATCTACCAC 840

Db 784 ACCAATGGAGCGGTGAATGGGAGGGTCACTCAAGGACAGCCCGGAGACATCTACMA - 842

QY 841 CAGACCTGGGCCAGATACTTGTGAACTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAG 900

Db 843 CAGACCTGGGCCAGATACTTGTGAACTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAG 902

QY 901 TTCTGGGAGTGACAGCTG - AAAATGAGCCCTTCTGCTGGGCTGTTGAGTGG 950

Db 903 TTCTGGGAGTGACAGCTGAAATAATGAGCCCTTCTGCTGGGCKGTTGAGTGG 953


```
RESULT 3
AL571401/c
LOCUS
DEFINITION AL571401 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI023Y009 3
prime, mRNA sequence.
ACCESSION AL571401
VERSION AL571401.1 GI:12928661
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 915)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI023Y009"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 208 a 234 c 270 g 201 t 2 others
ORIGIN
Query Match 39.6%; Score 901.6; DB 9; Length 915;
Best Local Similarity 99.9%; Pred. No. 4.9e-168;
Matches 901; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 835 TACCACAGACCTGGCCAGATACCTTTGTGAAGTCTCTGGATGCCTATGCTGAGCACAA 894
D 915 TACCACAGACCTGGCCAGATACCTTTGTGAAGTCTCTGGATGCCTATGCTGAGCACAA 856
QY 895 TTACAGTTCTGGCAGTGACAGCTGAAATGAGCTTCTGTGGCTGTGAGTGATAC 954
D 855 TTACAGTTCTGGCAGTGACAGCTGAAATGAGCTTCTGTGGCTGTGAGTGATAC 796
QY 955 CCCTCCAGTGCCTGGCTTCACCCCTGAACATCAGGAGACTTCATTGCCGTGACCTA 1014
D 795 CCCTCCAGTGCCTGGCTTCACCCCTGAACATCAGGAGACTTCATTGCCGTGACCTA 736
QY 1015 GGTCTACCTCGCCCAACAGTACTCACCACAATGTCGCCCTACTCATGCTGGATGACCA 1074
D 735 GGTCTACCTCGCCCAACAGTACTCACCACAATGTCGCCCTACTCATGCTGGATGACCA 676
QY 1075 CGCTGTGCTGCTGCCCACTGGCAAGGTGGTACTGACAGACCCAGAGCACTAAATAT 1134
D 675 CGCTGTGCTGCTGCCCACTGGCAAGGTGGTACTGACAGACCCAGAGCACTAAATAT 616
QY 1135 GTTCATGGCATTTGCTGATACATGGTACCTGGACTTCTGGCTCCAGCCAAAGCCACCTA 1194
D 615 GTTCATGGCATTTGCTGATACATGGTACCTGGACTTCTGGCTCCAGCCAAAGCCACCTA 556
QY 1195 GGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGC 1254
D 555 GGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGC 496
```

```
QY 1255 TCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGATGAGTACAGC 1314
D 495 TCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGATGAGTACAGC 436
QY 1315 CACAGCATCATCACGAACCTCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGC 1374
D 435 CACAGCATCATCACGAACCTCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGC 376
QY 1375 CTGAACCCCGAAGGAGGACCAATTTGGTGGTAACTTGTTCGACAGTCCCATCTGTA 1434
D 375 CTGAACCCCGAAGGAGGACCAATTTGGTGGTAACTTGTTCGACAGTCCCATCTGTA 316
QY 1435 GACATCACCAAGGACACAGTTTACAAACAGCCCATGTTCTACCATTTGGCCACTTCAGC 1494
D 315 GACATCACCAAGGACACAGTTTACAAACAGCCCATGTTCTACCATTTGGCCACTTCAGC 256
QY 1495 AAGTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAAGAAGACGACCTG 1554
D 255 AAGTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTCCAGTCAAGAAGACGACCTG 196
QY 1555 GACGAGTGGCACTGATGATCCCGATGGCTCTGCTGTGTGGTAAACCGCTCC 1614
D 195 GACGAGTGGCACTGATGATCCCGATGGCTCTGCTGTGTGGTAAACCGCTCC 136
QY 1615 TCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAATCTCA 1674
D 135 TCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAATCTCA 76
QY 1675 CCTGGTACTCCATTACACCTACCTGTGGCGTGGCCAGTGGAGCAGATACCTCAAGG 1734
D 75 CCTGGTACTCCATTACACCTACCTGTGGCGTGGCCAGTGGAGCAGATACCTCAAGG 16
QY 1735 AG 1736
D 15 AG 14
RESULT 4
AL568749/c
LOCUS
DEFINITION AL568749 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YL08 3 prime
, mRNA sequence.
ACCESSION AL568749
VERSION AL568749.1 GI:12923399
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 927)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 927
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE005YL08"
/clone_lib="LTI_FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
```

fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 199 a 266 c 249 g 211 t 2 others
ORIGIN

Query Match 39.3%; Score 896.2; DB 9; Length 927;
Best Local Similarity 99.0%; Pred. No. 5.6e-167;
Matches 920; Conservative 2; Mismatches 5; Indels 2; Gaps 2;

```
QY 1232 TTGCTCAGAGGCTGTGTGGGCTCCAAAGTTCTTGGAGCAGAGTGTGCGGCTAGGCTCCT 1291
Db 927 TTTCTCAGAGGCTKTGT-GGCTCCAAAGTTCTTGGAGCAGAGTGTGCGGCTAGGCTCCT 869

QY 1292 GGGATCGAGGATGCAGTACAGCCACAGATCATCACGAACCTCCTGTACCATGTGGTCG 1351
Db 868 GGGATCGAGGATGCAGTACAGCCACAGATCATCACGAACCTCCTGTACCATGTGGTCG 809

QY 1352 GCTGGACGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCCAATTGGGTGCGTAAC 1411
Db 808 GCTGGACGACT-GAACCCTTGCCCTGAACCCCGAAGGAGGACCCCAATTGGGTGCGTAAC 750

QY 1412 TTGTCGACAGTCCCATCATATTGTAGACATCACCAAGGACACGTTTACAAACAGCCCATGT 1471
Db 749 TTGTCGACAGTCCCATCATATTGTAGACATCACCAAGGACACGTTTACAAACAGCCCATGT 690

QY 1472 TCTACCACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGCTCCCAGAGAGTGGGCTGG 1531
Db 689 TCTACCACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGCTCCCAGAGAGTGGGCTGG 630

QY 1532 TTGCCAGTCAGAAAGAACGACCTTGGACGAGTGGCACTGATGCATCCCGATGGCTCTGCTG 1591
Db 629 TTGCCAGTCAGAAAGAACGACCTTGGACGAGTGGCACTGATGCATCCCGATGGCTCTGCTG 570

QY 1592 TTGTGGTGTGCTAAACCGCTCCTCTAAGGATGTCCCTTACCATCAAGGATCCCTGCTG 1651
Db 569 TTGTGGTGTGCTAAACCGCTCCTCTAAGGATGTCCCTTACCATCAAGGATCCCTGCTG 510

QY 1652 TGGGCTTCCTGGAGACAACTTCACTGGCTACTCCATTCACACCTACCTTGTGGCTGCGC 1711
Db 509 TGGGCTTCCTGGAGACAACTTCACTGGCTACTCCATTCACACCTACCTTGTGGCTGCGC 450

QY 1712 AGTGATGGAGCAGATCAAGAGGAGGCTGGCTCAGCTGGGCTGGGCTAAAGGGACAGAG 1771
Db 449 AGTGATGGAGCAGATCAAGAGGAGGCTGGCTCAGCTGGGCTGGGCTAAAGGGACAGAG 390

QY 1772 TCAGCTCACACGCTGTGTGACTAAAGAGGGCACAGAGGGCCAGTGTGAGCTTACAGC 1831
Db 389 TCAGCTCACACGCTGTGTGACTAAAGAGGGCACAGAGGGCCAGTGTGAGCTTACAGC 330

QY 1832 GACGTAAGCCAGGGCAATGGTTGGGTGACTCACTTCCCTCTAGGTGGTGGCCAGGG 1891
Db 329 GACGTAAGCCAGGGCAATGGTTGGGTGACTCACTTCCCTCTAGGTGGTGGCCAGGG 270

QY 1892 GCTGGAGGCCCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCCAGCCCCCATGCTTAT 1951
Db 269 GCTGGAGGCCCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCCAGCCCCCATGCTTAT 210

QY 1952 GTGAACATGCGTGTGTGCTTGGAACTGGGGCTGGGCTCCAGGCCCTAGGGTG 2011
Db 209 GTGAACATGCGTGTGTGCTTGGAACTGGGGCTGGGCTCCAGGCCCTAGGGTG 150

QY 2012 AGCTCACTGTCCGTACAAACACAGATCAGGGCTGAGGGTAAGGAAAGAGAGACTAGG 2071
Db 149 AGCTCACTGTCCGTACAAACACAGATCAGGGCTGAGGGTAAGGAAAGAGAGACTAGG 90

QY 2072 AAAGCTGGGCCCCAAACTGGAGACTGTTTGTCTTCTTGGAGATGCAGAACTGGGCCCCGT 2131
Db 89 AAAGCTGGGCCCCAAACTGGAGACTTTTTGTCTTCTTGGAGATGCAGAACTGGGCCCCGT 30

QY 2132 GGAGCAGCAGTGTACGATCAGGGCGGAA 2160
Db 29 GGAGCAGMAGTGTACGATCAGGGCGGAA 1
```

AL524792 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC008YG23 5
prime, mRNA sequence.
AL524792
AL524792.1 GI:12788285
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 940)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 940
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC008YG23"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

FEATURES
source

BASE COUNT 210 a 274 c 233 g 220 t 3 others
ORIGIN
Query Match 38.6%; Score 879.8; DB 9; Length 940;
Best Local Similarity 99.3%; Pred. No. 9.7e-164;
Matches 911; Conservative 3; Mismatches 0; Indels 3; Gaps 3;

```
QY 1 AGCTAAGGCAGGATACCTGCATCCTTGTGTTTGTAGTGATCCTCTATCCTTCAGAGAC 60
Db 26 AGCTAAGGCAGGATACCTGCATCCTTGTGTTTGTAGTGATCCTCTATCCTTCAGAGAC 85

QY 61 TCTGGAACCCCTGTGTCTCTCTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGT 120
Db 86 TCTGGAACCCCTGTGTCTCTCTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGT 145

QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGTAGTAGGTAAGCATCATGGCTGGCAGCCTC 180
Db 146 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGTAGTAGGTAAGCATCATGGCTGGCAGCCTC 205

QY 181 ACAGGATGTCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCCGCCCTGCATCCCT 240
Db 206 ACAGGATGTCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCCGCCCTGCATCCCT 265

QY 241 AAAAGCTTGGCTACAGCTCGGTGGTGTGTCTGTCAATGCCACATACCTGTGACTCCCTT 300
Db 266 AAAAGCTTGGCTACAGCTCGGTGGTGTGTCTGTCAATGCCACATACCTGTGACTCCCTT 325

QY 301 GACCCCGGACCTTCTCTGCTTGGTACCTTCAGCGCTATGAGATACACGACGAGTGGG 360
Db 326 GACCCCGGACCTTCTCTGCTTGGTACCTTCAGCGCTATGAGATACACGACGAGTGGG 385

QY 361 CGACGATGGAGCTGAGTATGGGCGCATCCAGGCTAATCACACGGGCACAGCCCTGCTA 420
```

```
Db 386 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 445
QY 421 CTGACCCCTGCAGCCAGAACAGAAGTTCAGAAAAGTGAAGGGATTGGAGGGGCCATGACA 480
Db 446 CTGACCCCTGCAGCCAGAACAGAAGTTCAGAAAAGTGAAGGGATTGGAGGGGCCATGACA 505
QY 481 GATGCTGCTCTCAACATCCTTGGCCCTGTACCCCCCTGCCCAAAATTTGCTACTTAAA 540
Db 506 GATGCTGCTCTCAACATCCTTGGCCCTGTACCCCCCTGCCCAAAATTTGCTACTTAAA 565
QY 541 TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 600
Db 566 TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 625
QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGATTCAGTTCCAGTACAAC 660
Db 626 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGATTCAGTTCCAGTACAAC 685
QY 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGTATTCACCGAGCCCTGCAG 720
Db 686 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGTATTCACCGAGCCCTGCAG 745
QY 721 TTGGCCCGAGCGTCCCGTTTCACTCCTTGGCCAGCCCTTGACATCACCCACTTGGCTCAAG 780
Db 746 TTGGCCCGAGCGTCCCGTTTCACTCCTTGGCCAGCCCTTGACATCACCCACTTGGCTCAAG 805
QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCA 840
Db 806 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCA 864
QY 841 CAGACCTGGCCAGATACCTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 900
Db 865 CAGACCTGGCCAGATACCTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 922
QY 901 TTCTGGGCAGTGACAGC 917
Db 923 TTCTGGGCAGTGACAGC 939
```

```
RESULT 6
BQ676149
LOCUS
DEFINITION BQ676149 920 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
AGENCOURT_8494833 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6294994
```

```
ACCESSION BQ676149
VERSION BQ676149.1 GI:21788828
```

```
KEYWORDS EST.
SOURCE human.
```

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM2500 row: 1 column: 11
High quality sequence start: 19
High quality sequence stop: 704.

FEATURES

source

1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6294994"
/clone_lib="NIH_MGC_112"

/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 214 a 279 c 225 g 202 t
ORIGIN

Query Match 37.9%; Score 864.6; DB 14; Length 920;
Best Local Similarity 98.2%; Pred. No. 9.8e-161;
Matches 895; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

```
QY 261 GGTGGTGTGTCTGCAATGCCACATACCTGTGACTCCTTTGACCCCCCGACCTTCCTGC 320
Db 10 GGTGGTGTGTCTGCAATGCCAC-TACTGTGACTCCTTTGACCCCCCGACCTTCCTGC 68
QY 321 CCTTGTACCTTTCAGCCGCTATGAGAGTACACGCGAGTGGCGACGGATGGAGCTGAGTAT 380
Db 69 CCTTGTACCTTTCAGCCGCTATGAGAGTACACGCGAGTGGCGACGGATGGAGCTGAGTAT 128
QY 381 GGGGCCCATCCAGGCTAATCAGCGGGCACAGGCGCTGCTACTGACCCCTGCAGCCAGAA 440
Db 129 GGGGCCCATCCAGGCTAATCAGCGGGCACAGGCGCTGCTACTGACCCCTGCAGCCAGAA 188
QY 441 GAAGTTCCAGAAAGTGAAGGGATTGGAGGGCCATGACAGATGCTGCTCTCAACAT 500
Db 189 GAAGTTCCAGAAAGTGAAGGGATTGGAGGGCCATGACAGATGCTGCTCTCAACAT 248
QY 501 CCTTGGCCCTGTCAACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAG 560
Db 249 CCTTGGCCCTGTCAACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAG 308
QY 561 AATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTA 620
Db 309 AATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTA 368
QY 621 CACCTATGCAGACACCCCTGTATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGA 680
Db 369 CACCTATGCAGACACCCCTGTATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGA 428
QY 681 TACCAAGCTCAAGATACCCCTGTATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCGTTTC 740
Db 429 TACCAAGCTCAAGATACCCCTGTATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCGTTTC 488
QY 741 ACTCCTTGCAGGCCCTTGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGG 800
Db 489 ACTCCTTGCAGGCCCTTGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGG 548
QY 801 GAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACAGACCTGGGCCAGATACTT 860
Db 549 GAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACAGACCTGGGCCAGATACTT 608
QY 861 TGTGAAGTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTGGGCAGTGACAGCTGA 920
Db 609 TGTGAAGTTCTCTGGATGCCCTATGCTGAGCACAAGTTACAGTTCTGGGCAGTGACAGCTGA 668
QY 921 AAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCC 980
Db 669 AAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCC 728
QY 981 TGAACATCAGCGAGACTTCAATTGCCGTGACCTAGGTCTTACCTCGCCCAACAGTACTCA 1040
Db 729 TGAACATCAGCGAGACTTCAATTGCCGTGACCTAGGTCTTACCTCGCCCAACAGTACTCA 788
QY 1041 CCACAATGTCCGCTACTCATGTGATGACCAACAGCTTGTGCTGCTCCCTGCTGGGCAAA 1100
Db 789 CCACAATGTCCGCTACTCATGTGATGACCAACAGCTTGTGCTGCTCCCTGCTGGGCAAA 848
```


QY 1101 GGTGGTACTGACAGACCCAGAACGACGCTAAATATGTTTCAT-GGCATTGCTGTACATTGGT 1159
|||||
Db 849 GGTGGTACTGACAGAACCCAAAGCAGCTTAATATGTTTCCTGGGCATTGCTGTACAATGGG 908
QY 1160 ACCTGGACTTT 1170
| | | | |
Db 909 AACCTGGACTT 919

RESULT 7
BG744064 916 bp mRNA linear EST 15-MAY-2001
LOCUS 602722846F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849307 5',
DEFINITION mRNA sequence.
ACCESSION BG744064
VERSION BG744064.1 GI:14054717
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 916)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1688 row: 1 column: 12
High quality sequence stop: 856.
Location/Qualifiers
1. 916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4849307"
/clone_lib="NIH_MGC_106"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 202 a 257 c 258 g 199 t
ORIGIN

Query Match 37.7%; Score 858.4; DB 12; Length 916;
Best Local Similarity 99.2%; Pred. NO. 1.6e-159;
Matches 873; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1156 TGGTACCTGGACTTTCTGGCTCCAGCAAGCCACCTAGGGGAGACACACCGCCTGTC 1215
|||||
Db 12 TGGTACCTGGACTTTCTGGCTCCAGCAAGCCACCTAGGGGAGACACACCGCCTGTC 71
QY 1216 CCCAACACCATGCTCTTTGCTCAGAGGCCCTGTGTGGGCTCCAAAGTCTGGGAGCAGAT 1275
|||||
Db 72 CCCAACACCATGCTCTTTGCTCAGAGGCCCTGTGTGGGCTCCAAAGTCTGGGAGCAGAT 131
QY 1276 GTGCGGCTAGGCTCTGGGATCGAGGGATGCAGTACGCCACAGCATCATCAGAACCTC 1335
|||||
Db 132 GTGCGGCTAGGCTCTGGGATCGAGGGATGCAGTACGCCACAGCATCATCAGAACCTC 191
QY 1336 CTGTACCATGTGTCGGCTGGACCGACTGGAACCTTGCCTGAACCCCGAAGGAGGACCC 1395
|||||

Db 192 CTGTACCATGTGGTCGGCTGGACCGACTGGAACTTGCCTGAACCCCGAAGGAGGACCC 251
QY 1396 AATTGGGTGCGTAACCTTTGTCGACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTT 1455
|||||
Db 252 AATTGGGTGCGTAACCTTTGTCGACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTT 311
QY 1456 TACAAAACAGCCCATGTTCTACCACTTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCC 1515
|||||
Db 312 TACAAAACAGCCCATGTTCTACCACTTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCC 371
QY 1516 CAGAGAGTGGGCTGGTTGCCAGTCAGAAGACGACCTGGACGAGTGGCAGTGCATGCAT 1575
|||||
Db 372 CAGAGAGTGGGCTGGTTGCCAGTCAGAAGACGACCTGGACGAGTGGCAGTGCATGCAT 431
QY 1576 CCCGATGGCTCTGCTGTTGTGGTCTGTGCTAAACCGCTCCTCTAAGGATGTGCCTCTTACC 1635
|||||
Db 432 CCCGATGGCTCTGCTGTTGTGGTCTGTGCTAAACCGCTCCTCTAAGGATGTGCCTCTTACC 491
QY 1636 ATCAAGGATCCTGCTGTGGGCTTCTGGAGACAATCTCACCTGGCTACTCCATTACACACC 1695
|||||
Db 492 ATCAAGGATCCTGCTGTGGGCTTCTGGAGACAATCTCACCTGGCTACTCCATTACACACC 551
QY 1696 TACCTGTGGCTCGCCAGTGTGGAGCAGATACTCAAGGAGGACACTGGGCTCAGCCTGGG 1755
|||||
Db 552 TACCTGTGGCTCGCCAGTGTGGAGCAGATACTCAAGGAGGACACTGGGCTCAGCCTGGG 611
QY 1756 CATTAAAGGACAGAGTCACTCACACGCTGTGTGCTAAAGAGGAGGACAGCAGGGCC 1815
|||||
Db 612 CATTAAAGGACAGAGTCACTCACACGCTGTGTGCTAAAGAGGAGGACAGCAGGGCC 671
QY 1816 AGTGTGAGCTTACAGCGACGCTAAGCCCCAGGGGCAATGGTTGGGTGACTCATTTCCTT 1875
|||||
Db 672 AGTGTGAGCTTACAGCGACGCTAAGCCCCAGGGGCAATGGTTGGGTGACTCATTTCCTT 731
QY 1876 CTAGGTGTGCCAGGGCTGGAGGGCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCC 1935
|||||
Db 732 CTAGGTGTGCCAGGGCTGGAGGGCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCC 791
QY 1936 CAGCCCCCATGCTTATGTGAACATCGGCTGTGTGCTGCTTTGGAACCTGGGCTGG 1995
|||||
Db 792 CAGCCCCCATGCTTATGTGAACATCGGCTGTGTGCTGCTTTGGAACCTGGGCTGG 851
QY 1996 GTCCAGGCTAGGGTGGAGCTCAC-TGTCCGTACAAACACA 2034
|||||
Db 852 GTCCAGGCTTAGGTGAGCTCACTTGTCCGTACAAACACA 891

RESULT 8
BG744733/c 957 bp mRNA linear EST 15-MAY-2001
LOCUS 602722846T1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849307 3',
DEFINITION mRNA sequence.
ACCESSION BG744733
VERSION BG744733.1 GI:14055386
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1688 row: 1 column: 12

Db 541 AAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACAGAGCTGGGCCAGATACCTTT 600
QY 862 GTGAAGTTCTCGATGCCTATGCTGAGCACAAGTTACAGTTCTGGCAGTGACAGCTGAA 921
Db 601 GTGAAGTTCTCGATGCCTATGCTGAGCACAAGTTACAGTTCTGGCAGTGACAGCTGAA 660
QY 922 AATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGCTTCACCCCT 981
Db 661 AATGAGCCTTCTGCTGGGCTGGTGAATGATACCCCTTCCAGTGCCTGGCTTCACCCCT 720
QY 982 GAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCTACCTCGCCCAACAGTACTAC 1041
Db 721 GAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCTACCTCGCCCAA-AGTACTAC 779
QY 1042 CACAATGTCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCTGCCCTGCGCAAG 1101
Db 780 CCAATGTCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCTGCCCTGCGCAAG 839
QY 1102 GTGGTACTGACAGACCCAGAGCAGCTAAATATGTTTCATGGCATGCTGTACATTGGTAC 1161
Db 840 G-GGGACTGACGACCCAGAGCAGCTAAATGTTTCATGGCATGCTGTGCGGAACCTTGGGTAC 898
QY 1162 CTGGACTTCTGG 1174
Db 899 CCGGACTTTTCGG 911

RESULT 11
BQ681028 BQ681028 838 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8208132 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6262714
DEFINITION 5', mRNA sequence.
ACCESSION BQ681028
VERSION BQ681028.1 GI:21793707
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 838)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCN2425 row: h column: 11
High quality sequence stop: 716.
Location/Qualifiers
1. 838

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6262714"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 191 a 262 c 202 g 183 t
ORIGIN

Query Match 36.6%; Score 835; DB 14; Length 838;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 835; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 262 GTGGTGTGTGTCTGCAATGCCACATACCTGTGACTCTTTGACCCCGGAGCTTTCTCTGCC 321
Db 1 GTGGTGTGTGTCTGCAATGCCACATACCTGTGACTCTTTGACCCCGGAGCTTTCTCTGCC 60
QY 322 CTGGTACCTTCAGCCGCTATGAGAGTACACCGAGTGGCGGAGTGGAGTGGAGTATG 381
Db 61 CTGGTACCTTCAGCCGCTATGAGAGTACACCGAGTGGCGGAGTGGAGTATG 120
QY 382 GGGCCCCATCCAGGCTAATCAGCGGACAGGCTGCTACTGACCTCGAGCCAGAACAG 441
Db 121 GGGCCCCATCCAGGCTAATCAGCGGACAGGCTGCTACTGACCTCGAGCCAGAACAG 180
QY 442 AAGTCCAGAAAGTGAAGGATTTGGAGGGGCGCATGACAGATGCTGCTCTCAACATC 501
Db 181 AAGTCCAGAAAGTGAAGGATTTGGAGGGGCGCATGACAGATGCTGCTCTCAACATC 240
QY 502 CTGGCCCTGTACCCCTGCCCCAAAATTTGCTACTTAAATCGTACTTTCTGGAAGAAGGA 561
Db 241 CTGGCCCTGTACCCCTGCCCCAAAATTTGCTACTTAAATCGTACTTTCTGGAAGAAGGA 300
QY 562 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 360
QY 622 ACCTATGCAGACACCCCTGATGATTTCCAGTTGGACAACTTCAGCCTCCAGAGGAAGAT 681
Db 361 ACCTATGCAGACACCCCTGATGATTTCCAGTTGGACAACTTCAGCCTCCAGAGGAAGAT 420
QY 682 ACCAAGCTCAAGATACCCCTGATTCCAGGAGCTTGGCCAGCTGGCCAGCTGGCTTTCA 741
Db 421 ACCAAGCTCAAGATACCCCTGATTCCAGGAGCTTGGCCAGCTGGCCAGCTGGCTTTCA 480
QY 742 CTCCTTGCAGCCCTGAGACATCACCCACTTGGTCAAGACCAATGGAGCGGTGAATGGG 801
Db 481 CTCCTTGCAGCCCTGAGACATCACCCACTTGGTCAAGACCAATGGAGCGGTGAATGGG 540
QY 802 AAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCCAGACCTGGGCGGAGATACCTT 861
Db 541 AAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCCAGACCTGGGCGGAGATACCTT 600
QY 862 GTGAAGTTCCTGGATGCTGATGAGCACAAGTACAGTTCTGGCAGTGACAGCTGAA 921
Db 601 GTGAAGTTCCTGGATGCTGATGAGCACAAGTACAGTTCTGGCAGTGACAGCTGAA 660
QY 922 AATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 981
Db 661 AATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 720
QY 982 GAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCTACCCCTCGCCCAACAGTACTAC 1041
Db 721 GAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCTACCCCTCGCCCAACAGTACTAC 780
QY 1042 CACAATGTCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCTGCCCTGCGG 1096
Db 781 CACAATGTCGCTACTCATGCTGGATGACCAACGCTTGTGCTGCTGCCCTGCGG 835

RESULT 12
BQ677036 BQ677036 916 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8209522 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6272665
DEFINITION 5', mRNA sequence.
ACCESSION BQ677036
VERSION BQ677036.1 GI:21789715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 916)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2451 row: 9 column: 02
High quality sequence stop: 535.
Location/Qualifiers
1. 916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6272665"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 210 a 276 c 235 g 193 t 2 others
ORIGIN
Query Match 36.6%; Score 833.4; DB 14; Length 916;
Best Local Similarity 97.1%; Pred. No. 1.4e-154;
Matches 879; Conservative 0; Mismatches 23; Indels 3; Gaps 3;
QY 262 GTGGGTGTCTGCAATGCCACATCTACTGTGACTCTTGTACCCCGGACCTTTCTGCCC 321
Db 1 GTGGGTGTCTGCAATGCCACATCTACTGTGACTCTTGTACCCCGGACCTTTCTGCCC 60
QY 322 CTTGGTACCTTCAGCGGCTATGAGAGTACACGCGAGTGGCGGACGGATGGAGCTGATG 381
Db 61 CTTGGTACCTTCAGCGGCTATGAGAGTACACGCGAGTGGCGGACGGATGGAGCTGATG 120
QY 382 GGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGACAG 441
Db 121 GGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGACAG 180
QY 442 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCGGATGACAGATGCTGCTCTCAACATC 501
Db 181 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCGGATGACAGATGCTGCTCTCAACATC 240
QY 502 CTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAATCGTACTTCTTGAAGAAGGA 561
Db 241 CTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAATCGTACTTCTTGAAGAAGGA 300
QY 562 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 360
QY 622 ACCATGACAGACACCCCTGATGATTTCCAGTTGCAAACTTCAGCCTCCAGAGGAAGAT 681
Db 361 ACCATGACAGACACCCCTGATGATTTCCAGTTGCAAACTTCAGCCTCCAGAGGAAGAT 420
QY 682 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCA 741
Db 421 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCA 480
QY 742 CTCCTTGCCAGCCCTGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGG 801
|||||

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 916)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2451 row: 9 column: 02
High quality sequence stop: 535.
Location/Qualifiers
1. 916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6272665"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 210 a 276 c 235 g 193 t 2 others
ORIGIN
Query Match 36.6%; Score 833.4; DB 14; Length 916;
Best Local Similarity 97.1%; Pred. No. 1.4e-154;
Matches 879; Conservative 0; Mismatches 23; Indels 3; Gaps 3;
QY 262 GTGGGTGTCTGCAATGCCACATCTACTGTGACTCTTGTACCCCGGACCTTTCTGCCC 321
Db 1 GTGGGTGTCTGCAATGCCACATCTACTGTGACTCTTGTACCCCGGACCTTTCTGCCC 60
QY 322 CTTGGTACCTTCAGCGGCTATGAGAGTACACGCGAGTGGCGGACGGATGGAGCTGATG 381
Db 61 CTTGGTACCTTCAGCGGCTATGAGAGTACACGCGAGTGGCGGACGGATGGAGCTGATG 120
QY 382 GGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGACAG 441
Db 121 GGGCCCATCCAGGCTAATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGACAG 180
QY 442 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCGGATGACAGATGCTGCTCTCAACATC 501
Db 181 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCGGATGACAGATGCTGCTCTCAACATC 240
QY 502 CTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAATCGTACTTCTTGAAGAAGGA 561
Db 241 CTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAATCGTACTTCTTGAAGAAGGA 300
QY 562 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 360
QY 622 ACCATGACAGACACCCCTGATGATTTCCAGTTGCAAACTTCAGCCTCCAGAGGAAGAT 681
Db 361 ACCATGACAGACACCCCTGATGATTTCCAGTTGCAAACTTCAGCCTCCAGAGGAAGAT 420
QY 682 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCA 741
Db 421 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCA 480
QY 742 CTCCTTGCCAGCCCTGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGG 801
|||||

```
Query Match      36.5%; Score 830.8; DB 13; Length 1017;
Best Local Similarity 98.2%; Pred. No. 4.5e-154;
Matches 862; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 1 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTAGTGGATCCTCTATCCTTCAGAGAC 60
Db 12 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTAGTGGATCCTCTATCCTTCAGAGAC 71

QY 61 TCTGGAACCCCTGTGGTCTTCTCTTCATCTAATGACCCTGAGGGGATGGAGTTTCAAGT 120
Db 72 TCTGGAACCCCTGTGGTCTTCTCTTCATCTAATGACCCTGAGGGGATGGAGTTTCAAGT 131

QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGAGTAGGGTAAGCATCATGGCTGGCAGCCCTC 180
Db 132 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGAGTAGGGTAAGCATCATGGCTGGCAGCCCTC 191

QY 181 ACAGGATTGCTTCTACTTTCAGGCAGTGTGGTGGGCATCAGGTGCCCGCCCTGCATCCCT 240
Db 192 ACAGGATTGCTTCTACTTTCAGGCAGTGTGGTGGGCATCAGGTGCCCGCCCTGCATCCCT 251

QY 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATCTGTGACTCCTTT 300
Db 252 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATCTGTGACTCCTTT 311

QY 301 GACCCCGGACCTTTCCTGCCCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGCAGTGG 360
Db 312 GACCCCGGACCTTTCCTGCCCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGCAGTGG 371

QY 361 CGACGGATGGAGTGTAGTATGGGGCCCTCCAGGCTAATCACACGGGCACAGGCCTGCTA 420
Db 372 CGACGGATGGAGTGTAGTATGGGGCCCTCCAGGCTAATCACACGGGCACAGGCCTGCTA 431

QY 421 CTGACCCCTGCAGCAGAACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 480
Db 432 CTGACCCCTGCAGCAGAACAGAAAGTTCCCGAAAGTGAAGGGATTGGAGGGGCCATGACA 491

QY 481 GATGCTGTGCTTCAACATCCTTGGCCCTGTCTACCCCTGCCCCAAATTTGCTACTTAA 540
Db 492 GATGCTGTGCTTCAACATCCTTGGCCCTGTCTACCCCTGCCCCAAATTTGCTACTTAA 551

QY 541 TCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCGGGTACCCATGGCCAGCTGT 600
Db 552 TCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCGGGTACCCATGGCCAGCTGT 611

QY 601 GACTTCTCCATCCGACCTACACCTATGCACACACCCCTGATGATTTCCAGTTGCACAAC 660
Db 612 GACTTCTCCATCCGACCTACACCTATGCACACACCCCTGATGATTTCCAGTTGCACAAC 671

QY 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
Db 672 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 731

QY 721 TTGGCCCGAGCGTCCGTTTCACTCCTTGCAGCCCTGGACATCACCCACTTGGCTCAAG 780
Db 732 TTGGCCCGAGCGTCCGTTTCACTCCTTGCAGCCCTGGACATCACCCACTTGGGCTCAG 791

QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCC-GGAGACATCTACCA 839
Db 792 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCAGGGGAGACATCTACCA 851

QY 840 CCAGACCTGGGCC---AGATACTTTGTGAAGTTCCTGG 874
Db 852 CCAGACCTGGGCCAGATAAATTTTGTGAAGTTCCTGG 889
```

```
RESULT 14
BQ897517
LOCUS
DEFINITION BQ897517 896 bp mRNA linear EST 16-AUG-2002
AGENCOURT_8095675 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6088015
5', mRNA sequence.
ACCESSION BQ897517
VERSION BQ897517.1 GI:22289531
```

```
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 896)
JOURNAL NIH-MGC http://mnc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2325 row: i column: 08
High quality sequence stop: 692.

FEATURES
Location/Qualifiers
1..896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6088015"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 211 a 271 c 214 g 197 t 3 others
ORIGIN
Query Match 36.4%; Score 829; DB 14; Length 896;
Best Local Similarity 98.0%; Pred. No. 1e-153;
Matches 838; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 262 GTGGTGTGTGTCTGCAATGCCACATCTGTGACTCTTGGACCCCTTCTCCTGCC 321
Db 1 GTGGTGTGTGTCTGCAATGCCACATCTGTGACTCTTGGACCCCTTCTCCTGCC 60

QY 322 CTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGCGGACGGATGGAGTGTAGTATG 381
Db 61 CTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGCGGACGGATGGAGTGTAGTATG 120

QY 382 GGGCCCATCCAGGCTAATCACACGGGCACAGGCCCTGCTACTGACCTTCAGCCAGAACAG 441
Db 121 GGGCCCATCCAGGCTAATCACACGGGCACAGGCCCTGCTACTGACCTTCAGCCAGAACAG 180

QY 442 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACAGATGCTGCTCTCAACATC 501
Db 181 AAGTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACAGATGCTGCTCTCAACATC 240

QY 502 CTTGGCCTGTACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 561
Db 241 CTTGGCCTGTACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 300

QY 562 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 360

QY 622 ACCATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGAT 681
Db 361 ACCATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGAT 420

QY 682 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCCGTTCA 741
Db 1
```


Search completed: February 22, 2003, 13:08:25
Job time : 2957 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 10:33:37 ; Search time 102 Seconds
(without alignments)
6852.121 Million cell updates/sec

Title: US-10-024-197-24
Perfect score: 2279
Sequence: 1 agctaaggcaggtacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/pctus_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | ID | Description |
|------------|--------|---------|--------|----|--------------------|
| 1 | 1601.6 | 70.3 | 1642 | 2 | US-08-713-928B-3 |
| 2 | 667 | 29.3 | 7620 | 1 | US-07-767-135-1 |
| 3 | 667 | 29.3 | 7620 | 1 | US-07-841-652-1 |
| 4 | 89.4 | 3.9 | 358 | 1 | US-07-841-652-2 |
| 5 | 77 | 3.4 | 89 | 1 | US-07-767-135-19 |
| 6 | 74 | 3.2 | 75 | 1 | US-07-767-135-2 |
| 7 | 49.6 | 2.2 | 7218 | 1 | US-08-232-463-14 |
| 8 | 42.2 | 1.9 | 289 | 4 | US-09-007-005-17 |
| 9 | 42.2 | 1.9 | 289 | 4 | US-09-244-796-17 |
| 10 | 41 | 1.8 | 41 | 1 | US-08-317-432A-44 |
| 11 | 40.8 | 1.8 | 7218 | 1 | US-08-232-463-14 |
| 12 | 39.4 | 1.7 | 41 | 1 | US-08-317-432A-46 |
| 13 | 39 | 1.7 | 41 | 1 | US-08-317-432A-48 |
| 14 | 37.4 | 1.6 | 41 | 1 | US-08-317-432A-50 |
| 15 | 36.6 | 1.6 | 379 | 3 | US-09-026-343-21 |
| 16 | 36.6 | 1.6 | 379 | 4 | US-09-362-871-21 |
| 17 | 36 | 1.6 | 53526 | 3 | US-08-658-136-2 |
| 18 | 36 | 1.6 | 53577 | 3 | US-08-658-136-1 |
| 19 | 35.2 | 1.5 | 11219 | 1 | US-07-642-734C-1 |
| 20 | 35.2 | 1.5 | 11219 | 3 | US-08-439-009A-1 |
| 21 | 34.8 | 1.5 | 1424 | 4 | US-09-326-186B-226 |
| 22 | 34.8 | 1.5 | 1428 | 5 | PCT-US94-09642-1 |
| 23 | 34.8 | 1.5 | 1957 | 1 | US-08-295-060-3 |
| 24 | 34.2 | 1.5 | 198 | 1 | US-08-330-108-16 |
| 25 | 34.2 | 1.5 | 198 | 5 | PCT-US92-10087-16 |
| 26 | 34 | 1.5 | 4143 | 4 | US-08-826-134-6 |
| 27 | 34 | 1.5 | 5350 | 4 | US-08-826-134-3 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 28 | 33.8 | 1.5 | 1952 | 1 | US-08-333-358-1 | Sequence 1, Appli |
| 29 | 33.8 | 1.5 | 1952 | 1 | US-08-463-694-1 | Sequence 1, Appli |
| 30 | 33.8 | 1.5 | 1952 | 1 | US-08-694-501-1 | Sequence 1, Appli |
| 31 | 33.8 | 1.5 | 2004 | 1 | US-08-471-033-18 | Sequence 18, Appl |
| 32 | 33.8 | 1.5 | 2004 | 1 | US-08-471-044-18 | Sequence 18, Appl |
| 33 | 33.8 | 1.5 | 2004 | 2 | US-08-463-483A-18 | Sequence 18, Appl |
| 34 | 33.8 | 1.5 | 2004 | 2 | US-08-471-046A-18 | Sequence 18, Appl |
| 35 | 33.8 | 1.5 | 2004 | 2 | US-08-470-566B-18 | Sequence 18, Appl |
| 36 | 33.8 | 1.5 | 2004 | 2 | US-08-469-334-18 | Sequence 18, Appl |
| 37 | 33.8 | 1.5 | 2004 | 3 | US-09-300-529-18 | Sequence 18, Appl |
| 38 | 33.8 | 1.5 | 2576 | 1 | US-08-471-033-35 | Sequence 35, Appl |
| 39 | 33.8 | 1.5 | 2576 | 2 | US-08-471-044-35 | Sequence 35, Appl |
| 40 | 33.8 | 1.5 | 2576 | 2 | US-08-463-483A-35 | Sequence 35, Appl |
| 41 | 33.8 | 1.5 | 2576 | 2 | US-08-471-046A-35 | Sequence 35, Appl |
| 42 | 33.8 | 1.5 | 2576 | 2 | US-08-470-566B-35 | Sequence 35, Appl |
| 43 | 33.8 | 1.5 | 2576 | 2 | US-08-469-334-35 | Sequence 35, Appl |
| 44 | 33.8 | 1.5 | 2576 | 3 | US-09-300-529-35 | Sequence 35, Appl |
| 45 | 33.8 | 1.5 | 2655 | 1 | US-08-471-033-17 | Sequence 17, Appl |

ALIGNMENTS

RESULT 1
US-08-713-928B-3
; Sequence 3, Application US/08713928B
; Patent No. 5929304
; GENERAL INFORMATION:
; APPLICANT: RADIN, DAVID N.
; APPLICANT: CRAMER, CAROLE L.
; APPLICANT: OISHI, KAREN K.
; APPLICANT: WEISSENBORN, DEBORAH L.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,928B
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/003,737
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7956-0011-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1642 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; US-08-713-928B-3

Query Match 70.3%; Score 1601.6; DB 2; Length 1642;
Best Local Similarity 99.8%; Pred. No. 0;


```
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7620 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: exon
LOCATION: 356..611
OTHER INFORMATION: /product= "Exon 1"
FEATURE:
NAME/KEY: intron
LOCATION: 612..979
OTHER INFORMATION: /function= "Intron 1"
FEATURE:
NAME/KEY: exon
LOCATION: 980..1067
OTHER INFORMATION: /product= "Exon 2"
FEATURE:
NAME/KEY: intron
LOCATION: 1068..1619
OTHER INFORMATION: /function= "Intron 2"
FEATURE:
NAME/KEY: exon
LOCATION: 1620..1811
OTHER INFORMATION: /product= "Exon 3"
FEATURE:
NAME/KEY: intron
LOCATION: 1812..1934
OTHER INFORMATION: /function= "Intron 3"
FEATURE:
NAME/KEY: exon
LOCATION: 1935..2081
OTHER INFORMATION: /product= "Exon 4"
FEATURE:
NAME/KEY: intron
LOCATION: 2082..3046
OTHER INFORMATION: /function= "Intron 4"
FEATURE:
NAME/KEY: exon
LOCATION: 3047..3180
OTHER INFORMATION: /product= "Exon 5"
FEATURE:
NAME/KEY: intron
LOCATION: 3181..3390
OTHER INFORMATION: /function= "Intron 5"
FEATURE:
NAME/KEY: exon
LOCATION: 3391..3563
OTHER INFORMATION: /product= "Exon 6"
FEATURE:
NAME/KEY: intron
LOCATION: 3564..4116
OTHER INFORMATION: /function= "Intron 6"
FEATURE:
NAME/KEY: exon
LOCATION: 4117..4354
OTHER INFORMATION: /product= "Exon 7"
FEATURE:
NAME/KEY: intron
LOCATION: 4355..5227
OTHER INFORMATION: /function= "Intron 7"
FEATURE:
NAME/KEY: exon
LOCATION: 5228..5452
OTHER INFORMATION: /product= "Exon 8"
FEATURE:
NAME/KEY: intron
```

```
LOCATION: 5453..5852
OTHER INFORMATION: /function= "Intron 8"
FEATURE:
NAME/KEY: exon
LOCATION: 5853..6016
OTHER INFORMATION: /product= "Exon 9"
FEATURE:
NAME/KEY: intron
LOCATION: 6017..6385
OTHER INFORMATION: /function= "Intron 9"
FEATURE:
NAME/KEY: exon
LOCATION: 6386..6502
OTHER INFORMATION: /product= "Exon 10"
FEATURE:
NAME/KEY: intron
LOCATION: 6503..6596
OTHER INFORMATION: /function= "Intron 10"
FEATURE:
NAME/KEY: exon
LOCATION: 6597..7245
OTHER INFORMATION: /function= "Exon 11"
US-07-767-135-1

Query Match 29.3%; Score 667; DB 1; Length 7620;
Best Local Similarity 89.1%; Pred. No. 2e-179;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;

QY 1489 TTCAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGGTGGTGGCCAGTCAGAAAGAAC 1548
Db 6381 TTCAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGGTGGTGGCCAGTCAGAAAGAAC 6440
QY 1549 GACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTTGTGCTGCTGCTAAAC 1608
Db 6441 GACCTGGACGCGAGTGGCACTGATGATCCCGATGGCTCTGCTGTTGTGCTGCTGCTAAAC 6500
QY 1609 C----- 1609
Db 6501 CGGTGAGGGCAATGGTGAGGTCTGGGAAGTGGGCTGAAGACAGACGCTTGGGGGCTTGGCA 6560
QY 1610 -----GCTCCTCTAAGGATGTGCTCTTAC 1634
Db 6561 GGATCACACTCTCAGCTTCTCCTCCCTGCTCCCTAGCTCTCTAAGGATGTGCTCTTAC 6620
QY 1635 CATCAAGGATCCTGCTGTGGGCTTCTCGGAGACAATCTCACCTGGCTACTCCATTACAC 1694
Db 6621 CATCAAGGATCCTGCTGTGGGCTTCTCGGAGACAATCTCACCTGGCTACTCCATTACAC 6680
QY 1695 CTACCTGTGGCGTCGCCAGTGATGGAGCAGATCTCAAGGAGGCACTGGGCTCAGCCTGG 1754
Db 6681 CTACCTGTGGCGTCGCCAGTGATGGAGCAGATCTCAAGGAGGCACTGGGCTCAGCCTGG 6740
QY 1755 GCATTAAAGGGACAGAGTCAGCTCACAGCTGTCTGTGACTAAAGAGGGCACAGAGGCG 1814
Db 6741 GCATTAAAGGGACAGAGTCAGCTCACAGCTGTCTGTGACTAAAGAGGGCACAGAGGCG 6800
QY 1815 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGCAATGGTTTGGTGACTCACTTTCCCC 1874
Db 6801 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGCAATGGTTTGGTGACTCACTTTCCCC 6860
QY 1875 TCTAGTGGTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCGCAGTGTCCCC 1934
Db 6861 TCTAGTGGTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCGCAGTGTCCCC 6920
QY 1935 CCAGCCCCCATGCTTATGTGAACATGGGCTGTGTGCTGCTTGTGTTGAAACTGGGCTG 1994
Db 6921 CCAGCCCCCATGCTTATGTGAACATGGGCTGTGTGCTGCTTGTGTTGAAACTGGGCTG 6980
QY 1995 GGTCCAGGCGCTAGGCTGAGCTCAGTCTCCGTACAAACACAAAGATCAGGGCTGAGGTAAG 2054
Db 6981 GGTCCAGGCGCTAGGCTGAGCTCAGTCTCCGTACAAACACAAAGATCAGGGCTGAGGTAAG 7040
QY 2055 GAAAGAAAGAGACTAGGAAAGCTGGGCCCCAAAACCTGGAGACTGTTTGTCTTTCTGGAGA 2114
```


Query Match 29.3%; Score 667; DB 1; Length 7620;
Best Local Similarity 89.1%; Pred. No. 2e-179;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;

QY 1489 TTCAGCAAGTTTCATCTCTGAGGGTCCAGAGAGTGGGGCTGGTGGTCCAGTCCAGAGAAAC 1548
|||||
Db 6381 TTCAGCAAGTTTCATCTCTGAGGGTCCAGAGAGTGGGGCTGGTGGTCCAGTCCAGAGAAAC 6440
|||||

QY 1549 GACCTGGACGCAGTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTTGTTGCTGCTAAAC 1608
|||||
Db 6441 GACCTGGACGCAGTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTTGTTGCTGCTAAAC 6500
|||||

QY 1609 C----- 1609
|

Db 6501 CGGTGAGGGCAATGGTGAGGTCTGGGAAGTGGGCTGAAGACAGCGTTGGGGCCCTTGGCA 6560
|

QY 1610 -----GCTCCTCTAAGGATGTGCTCTTAC 1634
|||||

Db 6561 GGATCACACTCTCAGCTTCTCCTCCTGCTCCCTAGCTCCTCTAAGGATGTGCTCTTAC 6620
|||||

QY 1635 CATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAACTCACTGCTGCTACTCCATTACACAC 1694
|||||

Db 6621 CATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAACTCACTGCTGCTACTCCATTACACAC 6680
|||||

QY 1695 CTACCTGTGGCTGCCAGTGTGATGAGCAGATATCAAGGAGGCTGGGCTCAGCTGG 1754
|||||

Db 6681 CTACCTGTGGCTGCCAGTGTGATGAGCAGATATCAAGGAGGCTGGGCTCAGCTGG 6740
|||||

QY 1755 GCATTAAGGGACAGAGTCACTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 1814
|||||

Db 6741 GCATTAAGGGACAGAGTCACTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 6800
|||||

QY 1815 CAGTGTGAGCTTACAGCAGCAGTAAAGCCAGGGCAATGTTTGGTGACTCAGCTTCCCC 1874
|||||

Db 6801 CAGTGTGAGCTTACAGCAGCAGTAAAGCCAGGGCAATGTTTGGTGACTCAGCTTCCCC 6860
|||||

QY 1875 TCTAGGTGTTCCAGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCTCAGTGTCCCC 1934
|||||

Db 6861 TCTAGGTGTTCCAGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCTCAGTGTCCCC 6920
|||||

QY 1935 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGCTGCTGCTTGGAACTGGGCTG 1994
|||||

Db 6921 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGCTGCTTGGAACTGGGCTG 6980
|||||

QY 1995 GGTCCAGGCCTAGGTGAGCTACTGCTCGTACAAACAAAGATCAGGGCTGAGGGTAAG 2054
|||||

Db 6981 GGTCCAGGCCTAGGTGAGCTACTGCTCGTACAAACAAAGATCAGGGCTGAGGGTAAG 7040
|||||

QY 2055 GAAAGAAGAGACTAGGAAAGCTGGGCCCCAAACTGGAGACTGTTCTTCTTCCCGAGA 2114
|||||

Db 7041 GAAAGAAGAGACTAGGAAAGCTGGGCCCCAAACTGGAGACTGTTCTTCTTCCCGAGA 7100
|||||

QY 2115 TGCAGAACTGGGCCCTGGAGCAGAGTGTGAGTCCAGTCCAGGGCGGAGCCCTTAAAGCAGCA 2174
|||||

Db 7101 TGCAGAACTGGGCCCTGGAGCAGAGTGTGAGTCCAGTCCAGGGCGGAGCCCTTAAAGCAGCA 7160
|||||

QY 2175 GCGGGTGTGCCAGGCACCCAGATGATTCCTATGSCACCCAGCCAGGAAAAATGGCAGCTC 2234
|||||

Db 7161 GCGGGTGTGCCAGGCACCCAGATGATTCCTATGSCACCCAGCCAGGAAAAATGGCAGCTC 7220
|||||

QY 2235 TTAAGGAGAAAAATGTTTGAGCCCA 2259
|||||

Db 7221 TTAAGGAGAAAAATGTTTGAGCCCA 7245
|||||

RESULT 4
US-07-841-652-2
; Sequence 2, Application US/07841652
; Patent No. 5266459
; GENERAL INFORMATION:
; APPLICANT: Beutler, Ernest
; TITLE OF INVENTION: GAUCHER'S DISEASE: DETECTION OF A NEW
; MUTATION IN INTRON 2 OF THE GLUCOCEREBROSIDASE GENE

NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10666 No. 5266459th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/841,652
FILING DATE: 19920224
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bingham, Douglas A
REGISTRATION NUMBER: 32,457
REFERENCE/DOCKET NUMBER: SCRO670P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-841-652-2

Query Match 3.9%; Score 89.4; DB 1; Length 358;
Best Local Similarity 98.9%; Pred. No. 7.2e-16;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 133 GAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGCTGGCAGCCTCACAGGATTGCTT 192
|||||
Db 1 GAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGCTGGCAGCCTCACAGGATTGCTT 60
|||||

QY 193 CTACTTCAGGCACTGCTGGGCATCAGGTG 223
|||||

Db 61 CTACTTCAGGCACTGCTGGGCATCAGATG 91
|||||

RESULT 5
US-07-767-135-19
; Sequence 19, Application US/07767135
; Patent No. 5234811
; GENERAL INFORMATION:
; APPLICANT: Beutler, Ernest
; APPLICANT: Sorge, Joseph A
; TITLE OF INVENTION: An Assay for a New Gaucher Disease
; TITLE OF INVENTION: Mutation
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 3366 No. 5234811th Torrey Pines Court, Suite 240
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/767,135
FILING DATE: 19910927
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bingham, Douglas A
REGISTRATION NUMBER: 32,457
REFERENCE/DOCKET NUMBER: SCR0422P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 89 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-767-135-19

Query Match 3.4%; Score 77; DB 1; Length 89;
Best Local Similarity 98.9%; Pred. No. 1.2e-12;
Matches 88; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 133 GAATGTCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTCACAGGATT-GCT 191
|||||
Db 1 GAATGTCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTCACAGGATTGGCT 60
|||||

QY 192 TCTACTTCAGGCAGTGTCTGGGGCATCAG 220
|||||
Db 61 TCTACTTCAGGCAGTGTCTGGGGCATCAG 89
|||||

RESULT 6
US-07-767-135-2
Sequence 2, Application US/07767135
Patent No. 5234811
GENERAL INFORMATION:
APPLICANT: Beutler, Ernest
APPLICANT: Sorge, Joseph A
TITLE OF INVENTION: An Assay for a New Gaucher Disease
TITLE OF INVENTION: Mutation
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 3366 No. 5234811th Torrey Pines Court, Suite 240
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/767,135
FILING DATE: 19910927
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bingham, Douglas A
REGISTRATION NUMBER: 32,457

REFERENCE/DOCKET NUMBER: SCR0422P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 57
OTHER INFORMATION: /note= "N is either G in a normal
OTHER INFORMATION: glucocerebrosidase gene or N is GG in a mutant
OTHER INFORMATION: glucocerebrosidase gene"
US-07-767-135-2

Query Match 3.2%; Score 74; DB 1; Length 75;
Best Local Similarity 98.7%; Pred. No. 7.7e-12;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 133 GAATGTCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTCACAGGATTGCTT 192
|||||
Db 1 GAATGTCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTCACAGGATTNCTT 60
|||||
QY 193 CTACTTCAGGCAGTG 207
|||||
Db 61 CTACTTCAGGCAGTG 75
|||||

RESULT 7
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-F1s
US-08-232-463-14

Query Match 2.2%; Score 49.6; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.00071;
Matches 4; Conservative 164; Mismatches 88; Indels 0; Gaps 0;

QY 2024 GTACAAACACAGATCAGGGCTGAGGCTAAGGAAAGAGAGACTAGGAAAGCTGGGCC 2083
Db 1438 GTACRR 1379
QY 2084 AAAACTGGAGACTGTTGCTTCTTCTGAGATGCAGAACTGGGCCGCTGGAGCAGTG 2143
Db 1378 RRR 1319
QY 2144 TCAGCATCAGGGCGGAGCCCTTAAAGCAGCAGCGGGTGTCGCCAGGCCAGATGATC 2203
Db 1318 RRR 1259
QY 2204 CTATGGCACCAGCCAGGAAATGCGAGCTCTTAAAGGAGAAATGTTGAGCCCAAAA 2263
Db 1258 RRR 1199
QY 2264 AAAAAAANAAAAA 2279
Db 1198 RRRRRRRRRRRRR 1183

RESULT 8
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.9%; Score 42.2; DB 4; Length 289;
Best Local Similarity 12.5%; Pred. No. 0.017;
Matches 33; Conservative 95; Mismatches 137; Indels 0; Gaps 0;

QY 2015 TCACTGTCCGTACAAACACAGATCAGGGCTGAGGCTAAGGAAAGAGAGACTAGGAAA 2074
Db 18 URARCRURARURURARCRARARURARCRARURARCRURARCRURARCRURGR 77

QY 2075 GCTGGGCCCAAACTGGAGACTGTTGCTTCTTCTGAGATGCAGAACTGGGCCCGTGA 2134
Db 78 NRSRNR 137
QY 2135 GCAGCAGTGTGAGCATCAGGGCGGAGCCCTTAAAGCAGCAGCGGGTGTCGCCAGGCC 2194
Db 138 NRSRNR 197
QY 2195 AGATGATTCCTATGCACAGCCAGGAAATGCGAGCTCTTAAAGGAGAGAAATGTTGA 2254
Db 198 NRSRNR 257
QY 2255 GCCCAAAAAAANAAAAA 2279
Db 258 CRUAAAAAANAAAAA 282

RESULT 9
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.9%; Score 42.2; DB 4; Length 289;
Best Local Similarity 12.5%; Pred. No. 0.017;
Matches 33; Conservative 95; Mismatches 137; Indels 0; Gaps 0;

QY 2015 TCACTGTCCGTACAAACACAGATCAGGGCTGAGGCTAAGGAAAGAGAGACTAGGAAA 2074
Db 18 URARCRURARURURARCRARARURARCRARURARCRURARCRURARCRURGR 77
QY 2075 GCTGGGCCCAAACTGGAGACTGTTGCTTCTTCTGAGATGCAGAACTGGGCCCGTGA 2134
Db 78 NRSRNR 137
QY 2135 GCAGCAGTGTGAGCATCAGGGCGGAGCCCTTAAAGCAGCAGCGGGTGTCGCCAGGCC 2194
Db 138 NRSRNR 197
QY 2195 AGATGATTCCTATGCACAGCCAGGAAATGCGAGCTCTTAAAGGAGAGAAATGTTGA 2254
Db 198 NRSRNR 257
QY 2255 GCCCAAAAAAANAAAAA 2279
Db 258 CRUAAAAAANAAAAA 282

Db 258 CRUAAAAAAAAAAAAAAAAAAAAA 282

RESULT 10

US-08-317-432A-44

; Sequence 44, Application US/08317432A
; Patent No. 5710028
; GENERAL INFORMATION:
; APPLICANT: Nurit Eyal and Nir Navot
; TITLE OF INVENTION: A method of quick screening and
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0
; SOFTWARE: converted to ASCII

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317.432A
; FILING DATE: 4-Oct-94

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/919,872

; FILING DATE: 27-Jul-92

; APPLICATION NUMBER: 08/084,505

; FILING DATE: 1-Jul-93

; ATTORNEY/AGENT INFORMATION:

; NAME: Friedmam, Mark M.

; REGISTRATION NUMBER: 33,883

; REFERENCE/DOCKET NUMBER: 128/7

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 972-3-5625553

; TELEFAX: 972-3-5625554

; TELEX:

; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 41

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-317-432A-44

Query Match 1.8%; Score 41; DB 1; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1533 TGCCAGTCAGAAGACGACCTGGACGCGAGTGGCGACTGATGC 1573

|||||

Db 1 TGCCAGTCAGAAGACGACCTGGACGCGAGTGGCGACTGATGC 41

RESULT 11

US-08-232-463-14

; Sequence 14, Application US/08232463
; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pt2gpt-F1s

US-08-232-463-14

Query Match 1.8%; Score 40.8; DB 1; Length 7218;

Best Local Similarity 5.4%; Pred. No. 0.23;

Matches 18; Conservative 176; Mismatches 138; Indels 0; Gaps 0;

QY 446 TCAGAAAGTGAAGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTG 505

|||||

Db 1031 TTCCGAGCTTGGCTGCAGGTCGAGGAGCTTCCGATTTTTTTTTTTTTTTTTT 1090

QY 506 CCTGTACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCG 565

|||||

Db 1091 YY 1150

QY 566 GATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCT 625

|||||

Db 1151 YY 1210

QY 626 ATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCCAGCTCCAGAGGAAGATACCA 685

|||||

Db 1211 YY 1270

QY 686 AGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCC 745

|||||

Db 1271 YY 1330

QY 746 TTGCCAGCCCTGGAGATCACCACCTTGGCTC 777

|||||

Db 1331 YY 1362

RESULT 12

US-08-317-432A-46

; Sequence 46, Application US/08317432A

; Patent No. 5710028

; GENERAL INFORMATION:

; APPLICANT: Nurit Eyal and Nir Navot

; TITLE OF INVENTION: A method of quick screening and

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
STREET: 2940 Birchtree lane
CITY: Silver Spring
STATE: Maryland
COUNTRY: United States of America
ZIP: 20906

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
OPERATING SYSTEM: Windows version 3.11
SOFTWARE: Word for Windows version 2.0
SOFTWARE: converted to ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,432A
FILING DATE: 4-Oct-94
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/919,872
FILING DATE: 27-Jul-92
APPLICATION NUMBER: 08/084,505
FILING DATE: 1-Jul-93

ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 128/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX:

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:
LENGTH: 41
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-317-432A-46

Query Match 1.7%; Score 39.4; DB 1; Length 41;
Best Local Similarity 97.6%; Pred. No. 0.039;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1533 TGCCAGTCAGAGAAGACGACCTGGACGAGTGGCAGTGC 1573
Db 1 TGCCAGTCAGAGAAGACGACCTGGACGAGTGGCAGTGC 41

RESULT 13

US-08-317-432A-48/c
Sequence 48, Application US/08317432A
Patent No. 5710028

GENERAL INFORMATION:
APPLICANT: Nurit Eyal and Nir Navot
TITLE OF INVENTION: A method of quick screening and
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
STREET: 2940 Birchtree lane
CITY: Silver Spring
STATE: Maryland
COUNTRY: United States of America
ZIP: 20906

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
OPERATING SYSTEM: Windows version 3.11
SOFTWARE: Word for Windows version 2.0
SOFTWARE: converted to ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,432A
FILING DATE: 4-Oct-94
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/919,872
FILING DATE: 27-Jul-92
APPLICATION NUMBER: 08/084,505
FILING DATE: 1-Jul-93

ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 128/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX:

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:
LENGTH: 41
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-317-432A-48

Query Match 1.7%; Score 39; DB 1; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1535 CCAGTCAGAGAAGACGACCTGGACGAGTGGCAGTGC 1573
Db 39 CCAGTCAGAGAAGACGACCTGGACGAGTGGCAGTGC 1

RESULT 14

US-08-317-432A-50/c
Sequence 50, Application US/08317432A
Patent No. 5710028

GENERAL INFORMATION:

APPLICANT: Nurit Eyal and Nir Navot
TITLE OF INVENTION: A method of quick screening and
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
STREET: 2940 Birchtree lane
CITY: Silver Spring
STATE: Maryland
COUNTRY: United States of America
ZIP: 20906

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
OPERATING SYSTEM: Windows version 3.11
SOFTWARE: Word for Windows version 2.0
SOFTWARE: converted to ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,432A
FILING DATE: 4-Oct-94
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/919,872
FILING DATE: 27-Jul-92
APPLICATION NUMBER: 08/084,505
FILING DATE: 1-Jul-93

ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 128/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX:

INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 41
TYPE: nucleic acid

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 12:19:34 ; Search time 176 Seconds
(without alignments)
7272.592 Million cell updates/sec

Title: US-10-024-197-24
Perfect score: 2279
Sequence: 1 agtaaggcagggtacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|---------------------|
| 1 | 1785.6 | 78.4 | 1792 | 10 | US-09-782-378A-16 |
| 2 | 1543 | 67.7 | 1551 | 10 | US-09-753-126-2 |
| 3 | 667 | 29.3 | 75270 | 9 | US-09-790-852-1 |
| 4 | 639.4 | 28.1 | 5769 | 10 | US-09-964-824A-284 |
| 5 | 225.4 | 9.9 | 362 | 10 | US-09-960-352-5649 |
| 6 | 130 | 5.7 | 135 | 9 | US-10-046-935-1845 |
| 7 | 130 | 5.7 | 135 | 9 | US-09-878-178-1845 |
| 8 | 39.6 | 1.7 | 376 | 10 | US-09-960-352-7227 |
| 9 | 38.4 | 1.7 | 477 | 10 | US-09-867-701-3268 |
| 10 | 38.4 | 1.7 | 2846 | 10 | US-09-833-381-1497 |
| 11 | 38.4 | 1.7 | 3612 | 10 | US-09-935-291A-10 |
| 12 | 38.4 | 1.7 | 3639 | 10 | US-09-833-381-1495 |
| 13 | 38.4 | 1.7 | 3919 | 10 | US-09-935-291A-8 |
| 14 | 36.6 | 1.6 | 379 | 12 | US-10-028-780-21 |
| 15 | 36.4 | 1.6 | 293 | 10 | US-09-880-107-1230 |
| 16 | 36 | 1.6 | 564 | 10 | US-09-833-381-1498 |
| 17 | 36 | 1.6 | 53522 | 9 | US-09-904-968A-1 |
| 18 | 35.8 | 1.6 | 964 | 10 | US-09-764-864-300 |
| 19 | 35.4 | 1.6 | 314 | 10 | US-09-867-701-10324 |

| | | | | | | | |
|---|----|------|-----|--------|----|---------------------|-------------------|
| c | 20 | 35.4 | 1.6 | 476 | 12 | US-10-001-843-87 | Sequence 87, Appl |
| | 21 | 35.2 | 1.5 | 509 | 10 | US-09-867-701-5694 | Sequence 5694, Ap |
| | 22 | 35 | 1.5 | 295 | 10 | US-09-867-701-8588 | Sequence 8588, Ap |
| | 23 | 34.8 | 1.5 | 1424 | 9 | US-09-954-531-366 | Sequence 366, App |
| | 24 | 34.8 | 1.5 | 1424 | 9 | US-09-441-411-21 | Sequence 21, Appl |
| | 25 | 34.8 | 1.5 | 1424 | 10 | US-09-962-436-556 | Sequence 556, App |
| c | 26 | 34.8 | 1.5 | 155074 | 9 | US-10-026-188-6 | Sequence 6, Appli |
| | 27 | 34.6 | 1.5 | 559 | 9 | US-09-883-152-27 | Sequence 27, Appl |
| | 28 | 34.6 | 1.5 | 623 | 9 | US-09-883-152-29 | Sequence 29, Appl |
| | 29 | 34.6 | 1.5 | 687 | 9 | US-09-774-639-106 | Sequence 106, App |
| c | 30 | 34.6 | 1.5 | 1665 | 10 | US-09-821-167-6 | Sequence 6, Appli |
| | 31 | 34.6 | 1.5 | 4308 | 9 | US-10-042-431-79 | Sequence 79, Appl |
| | 32 | 34.6 | 1.5 | 4308 | 9 | US-09-759-130B-449 | Sequence 449, App |
| c | 33 | 34.6 | 1.5 | 11188 | 10 | US-09-821-167-1 | Sequence 1, Appli |
| | 34 | 34.2 | 1.5 | 5175 | 12 | US-10-028-056-1 | Sequence 1, Appli |
| c | 35 | 34.2 | 1.5 | 15843 | 10 | US-09-764-869-2396 | Sequence 2396, Ap |
| c | 36 | 34.2 | 1.5 | 16774 | 10 | US-09-764-869-2395 | Sequence 2395, Ap |
| c | 37 | 34.2 | 1.5 | 16774 | 10 | US-09-764-869-2398 | Sequence 2398, Ap |
| c | 38 | 34 | 1.5 | 319 | 10 | US-09-867-701-2350 | Sequence 2350, Ap |
| | 39 | 34 | 1.5 | 1798 | 10 | US-09-925-299-230 | Sequence 230, App |
| c | 40 | 34 | 1.5 | 3945 | 10 | US-09-917-800A-1566 | Sequence 1566, Ap |
| | 41 | 34 | 1.5 | 53542 | 10 | US-09-801-574-61 | Sequence 61, Appl |
| c | 42 | 33.8 | 1.5 | 1267 | 12 | US-10-001-843-45 | Sequence 45, Appl |
| c | 43 | 33.6 | 1.5 | 398 | 10 | US-09-960-352-12846 | Sequence 12846, A |
| | 44 | 33.6 | 1.5 | 815 | 10 | US-09-894-998-1 | Sequence 1, Appli |
| | 45 | 33.6 | 1.5 | 1401 | 9 | US-09-738-626-3465 | Sequence 3465, Ap |

ALIGNMENTS

RESULT 1
US-09-782-378A-16
; Sequence 16, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-16

| | | | | | | | |
|-----------------------|-------|----------------------|--|------------|-----|--------|-------|
| Query Match | 78.4% | Score | 1785.6; | DB | 10; | Length | 1792; |
| Best Local Similarity | 99.8% | Pred. No. | 0; | | | | |
| Matches | 1788; | Conservative | 0; | Mismatches | 4; | Indels | 0; |
| | | Gaps | 0; | | | | |
| QY | 121 | CCITCCAGAGAGGAAATGTC | CCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTC | 180 | | | |
| Db | 1 | CCITCCAGAGAGGAAATGTC | CCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTC | 60 | | | |
| QY | 181 | ACAGGATTGCTTCTACTT | CAGGCAGTGTCTGTTGGGCATCAGGTGCCCCCCTGCATCCCT | 240 | | | |
| Db | 61 | ACAGGATTGCTTCTACTT | CAGGCAGTGTCTGTTGGGCATCAGGTGCCCCCCTGCATCCCT | 120 | | | |
| QY | 241 | AAAAGCTTCGGCTACAGCT | CGGTGGTGTGTGTCTGTCGAATGCCACATACGTGACTCCTTT | 300 | | | |
| Db | 121 | AAAAGCTTCGGCTACAGCT | CGGTGGTGTGTGTCTGTCGAATGCCACATACGTGACTCCTTT | 180 | | | |
| QY | 301 | GACCCCCCGACCTTCTCCT | GGTGGTACCTTACCCCGCTATGAGAGTACACGAGTGGG | 360 | | | |
| Db | 181 | GACCCCCCGACCTTCTCCT | GGTGGTACCTTACCCCGCTATGAGAGTACACGAGTGGG | 240 | | | |

QY 361 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 420
Db 241 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 300
QY 421 CTGACCCCTGCAGCAGAAACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 480
Db 301 CTGACCCCTGCAGCAGAAACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 360
QY 481 GATGCTGCTCTCAACATCCTTGGCCCTGTCAACCCCTGCCCCCTGCCCCAAATTTGCTACTTAA 540
Db 361 GATGCTGCTCTCAACATCCTTGGCCCTGTCAACCCCTGCCCCCTGCCCCAAATTTGCTACTTAA 420
QY 541 TCGTACTTCTCTGAAGAAAGAAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 600
Db 421 TCGTACTTCTCTGAAGAAAGAAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 480
QY 601 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGTATGATTTCCAGTTGCACAAC 660
Db 481 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGTATGATTTCCAGTTGCACAAC 540
QY 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTTCACCGAGCCCTGCAG 720
Db 541 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTTCACCGAGCACTGCAG 600
QY 721 TTGGCCCCAGCGTCCCGTTTCACTCCTTGGCCAGCCCTGGACATCACCCACTTGGCTCAAG 780
Db 601 TTGGCCCCAGCGTCCCGTTTCACTCCTTGGCCAGCCCTGGACATCACCCACTTGGCTCAAG 660
QY 781 ACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
Db 661 ACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 720
QY 841 CAGACCTGGGCCAGATACTTTGTGAAGTTCCCTGGATGCCCTATGCTGAGCACAAGTTACAG 900
Db 721 CAGACCTGGGCCAGATACTTTGTGAAGTTCCCTGGATGCCCTATGCTGAGCACAAGTTACAG 780
QY 901 TTCTGGGCAGTGACAGCTGAAATGAGCCCTTCTGCTGGCTGTTGAGTGGATACCCCTTC 960
Db 781 TTCTGGGCAGTGACAGCTGAAATGAGCCCTTCTGCTGGCTGTTGAGTGGATACCCCTTC 840
QY 961 CAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCT 1020
Db 841 CAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCT 900
QY 1021 ACCCTCGCCCAACAGTACTCACCACAATGTCCGCTACTCATGCTGGATGACCAAGCTTG 1080
Db 901 ACCCTCGCCCAACAGTACTCACCACAATGTCCGCTACTCATGCTGGATGACCAAGCTTG 960
QY 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAATATGTTTCA 1140
Db 961 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAATATGTTTCA 1020
QY 1141 GGCATTGCTGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1200
Db 1021 GGCATTGCTGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1080
QY 1201 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCCTGTGTGGCTCCAAG 1260
Db 1081 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCCTGTGTGGCTCCAAG 1140
QY 1261 TTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGAGTACAGCCACAGC 1320
Db 1141 TTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGAGTACAGCCACAGC 1200
QY 1321 ATCATCAGCAACCTCCTGTACCATGTGGTGGCTGACCGGACTGGAACCTTGGCCCTGAAC 1380
Db 1201 ATCATCAGCAACCTCCTGTACCATGTGGTGGCTGACCGGACTGGAACCTTGGCCCTGAAC 1260
QY 1381 CCCGAAGGAGGACCCCAATTGGGTGCGTAACCTTTGTCGACAGTCCCATCATTTGAGACATC 1440
Db 1261 CCCGAAGGAGGACCCCAATTGGGTGCGTAACCTTTGTCGACAGTCCCATCATTTGAGACATC 1320

QY 1441 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACCACCTTGGCCACTTCAGCAAGTTC 1500
Db 1321 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACCACCTTGGCCACTTCAGCAAGTTC 1380
QY 1501 ATTCTCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCCAGTCAAGAAGAACGACCTGGACGCA 1560
Db 1381 ATTCTCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCCAGTCAAGAAGAACGACCTGGACGCA 1440
QY 1561 CTGGCACTGATGATCCCGATCCCGATGGCTCTGCTGTTGGTCTGTTAAACCGCTCCTCTAAG 1620
Db 1441 CTGGCACTGATGATCCCGATCCCGATGGCTCTGCTGTTGGTCTGTTAAACCGCTCCTCTAAG 1500
QY 1621 GATGTCCTCTTACCATCAAGGATCCCTGCTGTTGGGCTTCTGGAGACAATCTCACCTGGC 1680
Db 1501 GATGTCCTCTTACCATCAAGGATCCCTGCTGTTGGGCTTCTGGAGACAATCTCACCTGGC 1560
QY 1681 TACTCCATTACACACCTACCTGTGGCTCGCCAGTGGAGGAGCAGATACCTCAAGGAGCAC 1740
Db 1561 TACTCCATTACACACCTACCTGTGGCTCGCCAGTGGAGGAGCAGATACCTCAAGGAGCAC 1620
QY 1741 TGGGCTCAGCCTGGGCATTAAAGGACAGAGTCAAGCTCACACGCTGTCTGTGACTAAAGA 1800
Db 1621 TGGGCTCAGCCTGGGCATTAAAGGACAGAGTCAAGCTCACACGCTGTCTGTGACTAAAGA 1680
QY 1801 GGGACACAGAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCCGAGGGGCAATGTTGGGT 1860
Db 1681 GGGACACAGAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCCGAGGGGCAATGTTGGGT 1740
QY 1861 GACTCACTTCCCTCTAGGTGGTGGCCAGGGGCTGGAGGCCCTAGAAAAAG 1912
Db 1741 GACTCACTTCCCTCTAGGTGGTGGCCAGGGGCTGGAGGCCCTAGAAAAAG 1792

RESULT 2

US-09-753-126-2
; Sequence 2, Application US/09753126
; Patent No. US20020127219A1
; GENERAL INFORMATION:
; APPLICANT: OKKELS, JENS SIGURD
; APPLICANT: JENSEN, ANNE DAM
; APPLICANT: HALKIER, TORBEN
; APPLICANT: JENSEN, RIKKE BOLDING
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME
; FILE REFERENCE: 31-000600US
; CURRENT APPLICATION NUMBER: US/09753,126
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PA 1999 01891
; PRIOR FILING DATE: 1999-12-30
; PRIOR APPLICATION NUMBER: 60/174,652
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PA 200 00865
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/210,984
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 60/211,124
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: PA 2000 01027
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/217,497
; PRIOR FILING DATE: 2000-07-11
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-753-126-2

Query Match 67.7%; Score 1543; DB 10; Length 1551;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 225
Db 1 ATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 60
QY 226 CGCCCTGCTGCTTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 285
Db 61 CGCCCTGCTGCTTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 120
QY 286 TACTGTGACTCCTTTGACCCCTCCGACCTTTCTGCTCCCTTGGTACCTTCAGCCGCTATGAG 345
Db 121 TACTGTGACTCCTTTGACCCCTCCGACCTTTCTGCTCCCTTGGTACCTTCAGCCGCTATGAG 180
QY 346 AGTACACGCTGGGCGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACG 405
Db 181 AGTACACGCTGGGCGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACG 240
QY 406 GGCACAGGCTGCTACTGACCTTCAGCCCTGCAGCCAGAACAGAAAGTTCAGAAAGTGAAGGGATT 465
Db 241 GGCACAGGCTGCTACTGACCTTCAGCCCTGCAGCCAGAACAGAAAGTTCAGAAAGTGAAGGGATT 300
QY 466 GGAGGGCCATGACAGATGCTGCTCTCAACATCCTTGGCCCTGTCAACCCCTGCCCAA 525
Db 301 GGAGGGCCATGACAGATGCTGCTCTCAACATCCTTGGCCCTGTCAACCCCTGCCCAA 360
QY 526 AATTGCTACTTAATCGTACTTCTGAGAAAGGAATCGGATATAACATATCCGGGTA 585
Db 361 AATTGCTACTTAATCGTACTTCTGAGAAAGGAATCGGATATAACATATCCGGGTA 420
QY 586 CCCATGGCAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGAT 645
Db 421 CCCATGGCAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGCACAACTTCCAGCTCCCGAGGAGGATACCAAGCTCAAGATACCCCTGATT 705
Db 481 TTCCAGTTGCACAACTTCCAGCTCCCGAGGAGGATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACCGAGCCTGCAGTTGGCCAGCGTCCGCTTCACTCTCTGCCAGCCCTGGACATCA 765
Db 541 CACCGAGCCTGCAGTTGGCCAGCGTCCGCTTCACTCTCTGCCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGTCACTCAAGGGACAGCCC 825
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGTCACTCAAGGGACAGCCC 660
QY 826 GGAGACATCTACCACCCAGACCTGGGCCAGATCTTGTGAAGTTCTCTGGATGCCCTATGCT 885
Db 661 GGAGACATCTACCACCCAGACCTGGGCCAGATCTTGTGAAGTTCTCTGGATGCCCTATGCT 720
QY 886 GAGCACAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 945
Db 721 GAGCACAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 780
QY 946 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTCAACATCAGCGAGCTTCAATTGCC 1005
Db 781 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTCAACATCAGCGAGCTTCAATTGCC 840
QY 1006 CGTGACCTAGGTCCCTACCTCGCCCAACAGTACTCACCAACAATGTCCGCTACTCATGCTG 1065
Db 841 CGTGACCTAGGTCCCTACCTCGCCCAACAGTACTCACCAACAATGTCCGCTACTCATGCTG 900
QY 1066 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
Db 901 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 1126 GCTAATATGTTTCAATGGCATGCTGTACATTTGTTACCTGGACTTCTGGCTCCAGCCAAA 1185
Db 961 GCTAATATGTTTCAATGGCATGCTGTACATTTGTTACCTGGACTTCTGGCTCCAGCCAAA 1020
QY 1186 GCCACCTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTTCTGCTCCAGAGGCC 1245
Db 1021 GCCACCTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTTCTGCTCCAGAGGCC 1080
QY 1246 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTCGGGCTAGGCTCCTGGGATCGAGGGATG 1305

Db 1081 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCTGGGATCGAGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGTGCTGGCTGGACCCACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGTGCTGGCTGGACCCACTGG 1200
QY 1366 AACCTTGCCCTGACCCCGAAGGAGGACCCAAATGGGTGCGTAACTTGTGCGACAGTCCC 1425
Db 1201 AACCTTGCCCTGACCCCGAAGGAGGACCCAAATGGGTGCGTAACTTGTGCGACAGTCCC 1260
QY 1426 ATCATTTAGACATCACCAAGGACACGTTTACAAAACAGCCCCATGTTCTACCACTTGGC 1485
Db 1261 ATCATTTAGACATCACCAAGGACACGTTTACAAAACAGCCCCATGTTCTACCACTTGGC 1320
QY 1486 CACTTCAGCAAGTTCACTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTGGCAGTCAAGAAG 1545
Db 1321 CATTTCAAGCAAGTTCACTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTGGCAGTCAAGAAG 1380
QY 1546 AACGACCTGGACGAGTGGCAGTGCATCCGATGGCTCTGCTGTGTGGTGGTGGTGGTA 1605
Db 1381 AACGACCTGGACGAGTGGCAGTGCATCCGATGGCTCTGCTGTGTGGTGGTGGTGGTA 1440
QY 1606 AACCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAG 1665
Db 1441 AACCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAG 1500
QY 1666 ACAATCTCACCTGGCTACTCCATTCACACCTACACCTGTGGCGTGGCCAGTGA 1716
Db 1501 ACAATCTCACCTGGCTACTCCATTCACACCTACACCTGTGGCGTGGCCAGTGA 1551

RESULT 3
US-09-790-852-1
; Sequence 1, Application US/09790852
; Publication No. US20030013178A1
; GENERAL INFORMATION:
; APPLICANT: Glins, E.
; APPLICANT: Sidransky, E.
; APPLICANT: Winfield, S.
; APPLICANT: Tayebi, N.
; APPLICANT: Martin, B.M.
; TITLE OF INVENTION: DNA sequence the glucocerebrosidase gene
; FILE REFERENCE: 14014.0296
; CURRENT APPLICATION NUMBER: US/09/790,852
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 75270
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-790-852-1

Query Match 29.3%; Score 667; DB 9; Length 75270;
Best Local Similarity 89.1%; Pred. No. 6.9e-177;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;

QY 1489 TTCAGCAAGTTCACTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTGGCCAGTCAAGAAG 1548
Db 38068 TTCAGCAAGTTCACTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTGGCCAGTCAAGAAG 38127
QY 1549 GACCTGGACGAGTGGCAGTGCATCCCGATGGCTCTGCTGTGTGGTGGTGGTAAAC 1608
Db 38128 GACCTGGACGAGTGGCAGTGCATCCCGATGGCTCTGCTGTGTGGTGGTGGTAAAC 38187
QY 1609 C----- 1609
Db 38188 CCGTGAGGGCAATGGTGAGGTCTGGGAAGTGGGCTGAAGACAGCGTTGGGGGCTTGGCA 38247
QY 1610 -----GCTCCTCTAAGGATGCTGCTCTTAC 1634
Db 38248 GGATCACACTCTCAGCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTAC 38307

QY 1635 CATCAAGGATCCTGCTGTGGGCTTCTCGAGACAAATCTCACCTGGCTACTCCATTCACAC 1694
Db 38308 CATCAAGGATCCTGCTGTGGGCTTCTCGAGACAAATCTCACCTGGCTACTCCATTCACAC 38367
QY 1695 CTACCTGTGGCTGCGCCAGTGATGGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGG 1754
Db 38368 CTACCTGTGGCTGCGCCAGTGATGGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGG 38427
QY 1755 GCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 1814
Db 38428 GCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 38487
QY 1815 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGGTTTGGTGACTCACTTTCCCC 1874
Db 38488 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGGTTTGGTGACTCACTTTCCCC 38547
QY 1875 TCTAGGTGGTCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCTCAGTGTCCCC 1934
Db 38548 TCTAGGTGGTCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCTCAGTGTCCCC 38607
QY 1935 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTTGTGCTTGGAAACTGGGCTG 1994
Db 38608 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTTGTGCTTGGAAACTGGGCTG 38667
QY 1995 GGTCCAGGCCCTAGGCTGAGCTCACTGCTCCGTACAAACACAAAGATCAGGGCTGAGGGTAAG 2054
Db 38668 GGTCCAGGCCCTAGGCTGAGCTCACTGCTCCGTACAAACACAAAGATCAGGGCTGAGGGTAAG 38727
QY 2055 GAAAAGAGAGACTAGGAAAGCTGGGCCCCAAAAGCTGGAGACTGTTTGTCTTCTCCTGGAGA 2114
Db 38728 GAAAAGAGAGACTAGGAAAGCTGGGCCCCAAAAGCTGGAGACTGTTTGTCTTCTCCTGGAGA 38787
QY 2115 TGCAGAACTGGGCCCTGGAGCAGCAGTGTGAGCTCAGGGCGGAGGCTTAAAGCAGCA 2174
Db 38788 TGCAGAACTGGGCCCTGGAGCAGCAGTGTGAGCTCAGGGCGGAGGCTTAAAGCAGCA 38847
QY 2175 GCGGGTGTCCCGCAGCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTC 2234
Db 38848 GCGGGTGTCCCGCAGCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTC 38907
QY 2235 TTAAAGGAGAAAAATGTTTGAGCCCA 2259
Db 38908 TTAAAGGAGAAAAATGTTTGAGCCCA 38932

RESULT 4

US-09-964-824A-284
; Sequence 284, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 284
; LENGTH: 5769
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(5769)
; OTHER INFORMATION: n=a,t,g or c

US-09-964-324A-284

Query Match 28.1%; Score 639.4; DB 10; Length 5769;
Best Local Similarity 87.6%; Pred. No. 9.9e-170;
Matches 758; Conservative 0; Mismatches 12; Indels 95; Gaps 2;
QY 1489 TTCAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTTCAGAGTCAAGAAAC 1548
Db 4792 TTCAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGGGCTGGTTCAGAGTCAAGAAAC 4851
QY 1549 GACCTGGACGCGAGTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTGTCGTCTAAAC 1608
Db 4852 GACCCGGACGCGAGTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTGTCGTCTAAAC 4911
QY 1609 C----- 1609
Db 4912 CGGTGAGGCAATGTTGAGGTCTGGGAAGTGGGCTGAAGACAGCGTGGGGGCTTGGCA 4971
QY 1610 -----GCTCCTCTAAGGATGTGCTCTTAC 1634
Db 4972 GGATCACACTCTCAGCTTCTCCTCCTGCTCCCTAGCTCCTCTAAGGATGTGCTCTTAC 5031
QY 1635 CATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAATCTCACCTGGCTACTCCATTCACAC 1694
Db 5032 CATCAAGGATCCTGCTGTGGGCTTCTCTGGAGACAATCTCACCTGGCTACTCCATTCACAC 5091
QY 1695 CTACCTGTGGCTGCGCCAGTGTGAGGAGCAGATCTCAAGGAGGCACTGGGCTCAGCCTGG 1754
Db 5092 CTACCTGTGGCTGCGCCAGTGTGAGGAGCAGATCTCAAGGAGGCACTGGGCTCAGCCTGG 5151
QY 1755 GCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 1814
Db 5152 GCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGAGGGC 5211
QY 1815 CAGTGTGAGCTTACAGCGAGCTAAGCCAGGGCAATGGTGTGGGTGACTACTCTCTCC 1874
Db 5212 CAGTGTGAGCTTACAGCGAGCTAAGCCAGGGCAATGGTGTGGGTGACTACTCTCTCC 5271
QY 1875 TCTAGGTGGTCCAGGGGCTGGAGGCCCTCTAGAAAAAGATCAGTAAGCCCTCAGTGTCC 1934
Db 5272 TCTAGGTGGTCCAGGGGCTGGAGGCCCTCTAGAAAAAGATCAGTAAGCCCTCAGTGTCC 5331
QY 1935 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTTGGAAACTGGGCTG 1994
Db 5332 CCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTTGGAAACTGGGCTG 5390
QY 1995 GGTCCAGGCCCTAGGGTGAAGTCACTGCTCCGTACAAACACAAAGATCAGGGCTGAGGGTAAG 2054
Db 5391 GGTCCAGGCCCTAGGGTGAAGTCACTGCTCCGTACAAACACAAAGATCAGGGCTGAGGGTAAG 5450
QY 2055 GAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTTGTCTTCTCTGGAGA 2114
Db 5451 GAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTTGTCTTCTCTGGAGA 5510
QY 2115 TGCAGAACTGGGCCCTGGAGCAGCAGTGTGAGCTCAGGCGGGAAGCCCTTAAAGCAGCA 2174
Db 5511 TNNNNNNCTGGGCCCTGGAGCAGCAGTGTGAGCTCAGCATCAGGGCGGGAAGCCCTTAAAGCAGCA 5570
QY 2175 GCGGGTGTCCCGCAGGCCACCCAGATGATTCCTATGGCACCAGGAAAAATGGCAGCTC 2234
Db 5571 GCGGGTGTCCCGCAGGCCACCCAGATGATTCCTATGGCACCAGGAAAAATGGCAGCTC 5630
QY 2235 TTAAAGGAGAAAAATGTTTGAGCCCA 2259
Db 5631 TTAAAGGAGAAAAATGTTTGAGCCCA 5655

RESULT 5

US-09-960-352-5649
; Sequence 5649, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.

```

; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5649
; LENGTH: 362
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 24-LIB34-044-Q1-E1-F7
; DS-09-960-352-5649

```

| Query Match | 9.9%; | Score 225.4; | DB 10; | Length 362; |
|-----------------------|--------------|--|----------------|-------------|
| Best Local Similarity | 78.1%; | Pred. NO. 1.1e-53; | | |
| Matches 271; | Conservative | 0; | Mismatches 76; | Indels 0; |
| Gaps | 0; | | | |
| Qy | 712 | GGCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCCTTGCCAGCCCTCGGACATCACCCACT | 771 | |
| Db | 16 | GCAGGCCACCGCCAAACCGCTTGACTCACTCTTCGCCAGTCCCTGGACATCACCCACT | 75 | |
| Qy | 772 | TGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGGTCACTCAAGGGACAGCCCCGGAGAC | 831 | |
| Db | 76 | TGGCTCAAGACTAATGGGGCTGTGAATGGGAGGGGACACTCAAGGGTCACGCCAAGGGAT | 135 | |
| Qy | 832 | ATCTACCACCGAGACCTGGGCCAGATACTTTGTGAAGTTCCTGGATGCCTATGCTGAGCAC | 891 | |
| Db | 136 | CTCTACCACAAGACCTGGGCCAGATACTTTGTCAAGTTCCTGGATGCCTACGCCGAGCAT | 195 | |
| Qy | 892 | AAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGCTGTTGAGTGGA | 951 | |
| Db | 196 | AAGTTACAGTTCTGGGCAGTGACAGCCGAGAACGATCCTACTGCATGGCTCCTTACCGGG | 255 | |
| Qy | 952 | TACCCCTTCCAGTGCCTGGGCCTTCAACCCCTGAACATCAGCGAGACTTCATTGCCCGTGAC | 1011 | |
| Db | 256 | TACCCCTTCCAAATGCCTGGGTTTCACTCCTGAACATCAGCTAGACTTCATGCCCGTGAC | 315 | |
| Qy | 1012 | CTAGGTCCCTACCCCTCGCCCAACAGTACTCACCAACAATGTCCGCCCTACT | 1058 | |
| Db | 316 | CTGGGGTCTCAACCGTTCGCCAAGACGACACAGCAGCGCTACGTCCCGAT | 362 | |

```

RESULT 6
US-10-046-935-1845/c
; Sequence 1845, Application US/10046935
; Patent No. US2020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1845
; LENGTH: 135
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 46, 52, 53
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1845

```

Query Match 5.7% Score 130; DB 9; Length 135;

[illegible]

```

RESULT 7
US-09-878-178-1845/c
; Sequence 1845, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1845
; LENGTH: 135
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(135)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1845

```

| | Query Match | 5.7%; | Score 130; | DB 9; | Length 135; |
|----|-----------------------|---|--------------------|-----------|-------------|
| | Best Local Similarity | 97.7%; | Pred. No. 4.9e-27; | | |
| | Matches 130; | Conservative 0; | Mismatches 3; | Indels 0; | Gaps 0; |
| QY | 1880 | GTGGTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCCAGC | 1939 | | |
| Db | 135 | GTGGTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCCCAGTGTCCCCCAGC | 76 | | |
| QY | 1940 | CCCCATGCTTATGTGAACATGCCTGTGTGCTGCTTGTGCTTTGGAAACTGGGCCTGGGTCC | 1999 | | |
| Db | 75 | CCCCATGCTTATGTGAACATGCNNTGTGTNCTGCTTGTCTTTGGAAACTGGGCCTGGGTCC | 16 | | |
| QY | 2000 | AGGCCTAGGGTGA | 2012 | | |
| Db | 15 | AGGCCTAGGGTGA | 3 | | |

RESULT 8
US-09-960-352-7227/c
; Sequence 7227, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: MUSCLE AND FAT D
; FILE REFERENCE: 16511.006/37-21(10298
; CURRENT APPLICATION NUMBER: US/09/960
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112


```
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-028-780-21

Query Match      1.6%; Score 36.6; DB 12; Length 379;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 90; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1504 CCTGAGGGCTCCAGAGAGTGGGGCTGTTGCCAGTACAGAAAGACGACCTGGACGGCAGTG 1563
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13 CCTGAAGGCTACAAGAAACCCGGAGCTACTGCTAGACTCCAGAAAGATGGTGTCAATCAA 72

QY 1564 GCACTGATGCATCCCGATGGCTCTGCTGTGTGGTCGTAAACCGCTCCTCTAAAGGAT 1623
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 73 AAAGACAAGAACTCCCTGGGAGCAATCTGCAACAGGTAGCCAATCTGAATCTTAAGGAC 132

QY 1624 GTGCCTCTTACCATCAAGGATCCTGTGTGGGCTTCCTGGAGACAATCTCACCTGGCTAC 1683
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 133 CTCTCATATACCTTAAGGATTATGTTTTTAAAGAGCTTCAAAGAGACTGGCCTGGNTAC 192

RESULT 15
US-09-880-107-1230/c
; Sequence 1230, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1230
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA460665
US-09-880-107-1230

Query Match      1.6%; Score 36.4; DB 10; Length 293;
Best Local Similarity 61.7%; Pred. No. 1.8;
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2185 CCAGGCACCCAGATGATTCCTATGCGCACCAGCCAGGAAATGGCAGCTCTTAAAGGAGA 2244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 CCACCTGCCCCCTGCTTCTTAAGAGATGGGGGAGAAACTTCCAAATGTTTATAAGA 35

QY 2245 AAATGTTTACCCCAAAAAAAAAAAAAAAAAAAAA 2278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 AAGTCTGTACCACAAAAAAAAAAAAAAAAAAAAAA 1
```

Search completed: February 22, 2003, 14:38:46
Job time : 330 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 22, 2003, 08:17:37 ; Search time 5848 Seconds
(without alignments)
11341.535 Million cell updates/sec
Title: US-10-024-197-24
Perfect score: 2279
Sequence: 1 agctaaggcaggtacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|-----------|--------------------|
| 1 | 2279 | 100.0 | 2279 | 9 | BC003356 | BC003356 Homo sapi |
| 2 | 2245.4 | 98.5 | 2259 | 9 | HUMGCA | D13286 Human mRNA |
| 3 | 2238 | 98.2 | 2275 | 9 | HUMGCBPRC | M19285 Human gluco |
| 4 | 2220 | 97.4 | 2587 | 9 | HUMGCB | M16328 Human gluco |
| 5 | 1785.6 | 78.4 | 1792 | 9 | HUMGCB | K02920 Human lysos |
| 6 | 1640.2 | 72.0 | 1661 | 6 | I09351 | I09351 Sequence 1 |
| 7 | 1587 | 69.6 | 1611 | 6 | AX147658 | AX147658 Sequence |
| 8 | 1547.8 | 67.9 | 1551 | 6 | AX147656 | AX147656 Sequence |
| 9 | 1544.2 | 67.8 | 1823 | 9 | HUMGCA | D13287 Human mRNA |
| 10 | 1544 | 67.7 | 1593 | 6 | AX299779 | AX299779 Sequence |
| 11 | 1543 | 67.7 | 1551 | 6 | AX191972 | AX191972 Sequence |
| 12 | 1543 | 67.7 | 1551 | 6 | AX349364 | AX349364 Sequence |
| 13 | 1529.6 | 67.1 | 1592 | 6 | AX299777 | AX299777 Sequence |
| 14 | 1343 | 58.9 | 1900 | 9 | BC000349 | BC000349 Homo sapi |
| 15 | 1287 | 56.5 | 2711 | 9 | AK090514 | AK090514 Homo sapi |
| 16 | 1172.4 | 51.4 | 1964 | 10 | BC006663 | BC006663 Mus muscu |
| 17 | 1167 | 51.2 | 1684 | 10 | MUSGCB | M24119 Mouse gluco |
| 18 | 667 | 29.3 | 8850 | 9 | HUMGCB1 | J03059 Human gluco |
| 19 | 667 | 29.3 | 75270 | 9 | AF023268 | AF023268 Homo sapi |
| 20 | 667 | 29.3 | 178229 | 2 | AL713999 | AL713999 Homo sapi |
| 21 | 660.6 | 29.0 | 5420 | 9 | AF267177 | AF267177 Homo sapi |
| 22 | 659 | 28.9 | 7128 | 9 | AF285236 | AF285236 Pan trogl |
| 23 | 652.6 | 28.6 | 5498 | 9 | AF272641 | AF272641 Gorilla g |
| 24 | 639.4 | 28.1 | 5769 | 6 | AX335078 | AX335078 Sequence |
| 25 | 639.4 | 28.1 | 5769 | 9 | HUMGCB2 | J03060 Human gluco |
| 26 | 620.6 | 27.2 | 5494 | 9 | AF272642 | AF272642 Pan trogl |
| 27 | 529.4 | 23.2 | 543 | 11 | G06424 | G06424 human STS W |
| 28 | 180.2 | 7.9 | 16507 | 10 | AY115108 | AY115108 Mus muscu |
| 29 | 180.2 | 7.9 | 210496 | 10 | AC104327 | AC104327 Mus muscu |
| 30 | 171 | 7.5 | 188 | 9 | HUMGCB1 | M22212 Human gluco |
| 31 | 163.4 | 7.2 | 184 | 9 | HUMGCBGD | M20282 Human lysos |
| 32 | 160.4 | 7.0 | 162 | 9 | HUMGCBDA | M20248 Human gluco |
| 33 | 141.8 | 6.2 | 345 | 4 | AF134220 | AF134220 Equus cab |
| 34 | 130.4 | 5.7 | 138 | 6 | AX341598 | AX341598 Sequence |
| 35 | 127.2 | 5.6 | 403 | 9 | HUMGCBPRA | M18916 Human gluco |
| 36 | 117.6 | 5.2 | 403 | 9 | HUMGCBPRB | M18917 Human gluco |
| 37 | 108.4 | 4.8 | 1949 | 8 | AF352032 | AF352032 Phytophth |
| 38 | 102.4 | 4.5 | 11594 | 1 | AE012205 | AE012205 Xanthomon |
| 39 | 98.8 | 4.3 | 1785 | 3 | AY060718 | AY060718 Drosophil |
| 40 | 92.2 | 4.0 | 9661 | 1 | AE013177 | AE013177 Thermoana |
| 41 | 91.8 | 4.0 | 10007 | 1 | AE011746 | AE011746 Xanthomon |
| 42 | 84.4 | 3.7 | 95 | 9 | S44217 | S44217 beta-glucos |
| 43 | 82.8 | 3.6 | 95 | 9 | S44219 | S44219 beta-glucos |
| 44 | 80.6 | 3.5 | 36958 | 3 | CEFL1E6 | Z81058 Caenorhabdi |
| 45 | 79.6 | 3.5 | 95 | 9 | S44552 | S44552 beta-glucos |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|------|--------|-----------------|
| RESULT 1 | BC003356 | 2279 bp | mRNA | linear | PRI 12-JUL-2001 |
| LOCUS | Homo sapiens | | | | |
| DEFINITION | Homo sapiens, glucosidase, beta; acid (includes glucosylceramidase), clone MGC:5191 IMAGE:2899915, mRNA, complete cds. | | | | |
| ACCESSION | BC003356 | | | | |
| VERSION | BC003356.1 | GI:13097170 | | | |
| KEYWORDS | MGC. | | | | |
| SOURCE | Homo sapiens. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | 1 (bases 1 to 2279) | | | | |
| AUTHORS | Strausberg,R. | | | | |

Db 1321 ATCATCACGAACCTCCTGTACCATGTGGTCGGCTGGAGCCGACTGGAACCTTGCCCTGAAC 1380
QY 1381 CCCAAGGAGGACCCAAATTGGTGCCTAACTTTGTGCACAGTCCCATCATTTGTAGACATC 1440
Db 1381 CCCAAGGAGGACCCAAATTGGTGCCTAACTTTGTGCACAGTCCCATCATTTGTAGACATC 1440
QY 1441 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTC 1500
Db 1441 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTC 1500
QY 1501 ATTCCTGAGGGCTCCAGAGAGTGGGCTGGTTGCCAGTCAGAAGACGACCTGGACGCA 1560
Db 1501 ATTCCTGAGGGCTCCAGAGAGTGGGCTGGTTGCCAGTCAGAAGACGACCTGGACGCA 1560
QY 1561 GTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTGGTGGTAAACCGCTCCTCTAAG 1620
Db 1561 GTGGCACTGATGCATCCCGATGGCTCTGCTGTTGTGGTGGTAAACCGCTCCTCTAAG 1620
QY 1621 GATGTCCTCTTACCATCAAGGATCCCTGCTGTGGGCTTCCTGGAGACAATCTCACCTGGC 1680
Db 1621 GATGTCCTCTTACCATCAAGGATCCCTGCTGTGGGCTTCCTGGAGACAATCTCACCTGGC 1680
QY 1681 TACTCCATTACACCTACCTGTGGCGTCGCCAGTGATGGAGCAGATACTCAAGGAGGCAC 1740
Db 1681 TACTCCATTACACCTACCTGTGGCGTCGCCAGTGATGGAGCAGATACTCAAGGAGGCAC 1740
QY 1741 TGGGCTCAGCCTGGGCATTAAAGGGACAGAGTCAGCTCACACGCTGCTGTGACTAAAGA 1800
Db 1741 TGGGCTCAGCCTGGGCATTAAAGGGACAGAGTCAGCTCACACGCTGCTGTGACTAAAGA 1800
QY 1801 GGGCACAGCAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCCGAGGGCAATGGTTGGGT 1860
Db 1801 GGGCACAGCAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCCGAGGGCAATGGTTGGGT 1860
QY 1861 GACTCACTTTCCCTCTAGGTGGTGGCCAGGGGCTGAGGGCCCTAGAAAAAGATCAGTAA 1920
Db 1861 GACTCACTTTCCCTCTAGGTGGTGGCCAGGGGCTGAGGGCCCTAGAAAAAGATCAGTAA 1920
QY 1921 GCCCAGTGTCCCCAGCCGCCATGCTTATGTGAACATGCGTGTGCTGCTTGGCTTT 1980
Db 1921 GCCCAGTGTCCCCAGCCGCCATGCTTATGTGAACATGCGTGTGCTGCTTGGCTTT 1980
QY 1981 GGAACTGGCCCTGGGTCCAGGCCCTAGGGTGAGCTCACTGTCGTAACAACACAGATCA 2040
Db 1981 GGAACTGGCCCTGGGTCCAGGCCCTAGGGTGAGCTCACTGTCGTAACAACACAGATCA 2040
QY 2041 GGGCTGAGGGTAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTT 2100
Db 2041 GGGCTGAGGGTAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTT 2100
QY 2101 GTCTTCTCGAGATGCAGAACTGGGCCCGTGGAGCAGCAGTGTCAAGCATCAGGGCGGAA 2160
Db 2101 GTCTTCTCGAGATGCAGAACTGGGCCCGTGGAGCAGCAGTGTCAAGCATCAGGGCGGAA 2160
QY 2161 GCCTTAAAGCAGCAGCGGGTGTGCCAGGCCACCCAGATGATTCCTATGGCACCAGCCAGG 2220
Db 2161 GCCTTAAAGCAGCAGCGGGTGTGCCAGGCCACCCAGATGATTCCTATGGCACCAGCCAGG 2220
QY 2221 AAAAATGGCAGCTCTTAAAGGAGAGAAATGTTTGGCCCAAAAAAAGAAAAA 2279
Db 2221 AAAAATGGCAGCTCTTAAAGGAGAGAAATGTTTGGCCCAAAAAAAGAAAAA 2279

RESULT 2
HUMGCA
LOCUS HUMGCA 2259 bp mRNA linear PRI 01-FEB-2000
DEFINITION Human mRNA for glucocerebrosidase, complete cds.
ACCESSION D13286
VERSION D13286.1 GI:496368
KEYWORDS glucocerebrosidase.
SOURCE Homo sapiens (library: lambda hgc18) cell-line HL-60 cDNA to mRNA, clone pGC18A.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT On Jun 5, 1994 this sequence version replaced gi:441042.

FEATURES
source

CDS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2259)
Imai,K., Nakamura,M., Yamada,M., Asano,A., Yokoyama,S., Tsuji,S.
and Ginns,E.I.
A novel transcript from a pseudogene for human glucocerebrosidase
in non-Gaucher disease cells
Gene 136 (1-2), 365-368 (1993)
94124033
2 (bases 1 to 2259)
Nakamura,M.
Direct Submission
Submitted (21-SEP-1992) Masahiko Nakamura, Osaka University,
Institute for Protein Research; 3-2 Yamadaoka, Suita, Osaka 565,
Japan (E-mail:masahiko@protein.osaka-u.ac.jp, Tel:06-879-8628,
Fax:06-879-8629)
On Jun 5, 1994 this sequence version replaced gi:441042.

Location/Qualifiers
1..2259

/organism="Homo sapiens"
/db_xref="taxon:9606"

/cell_line="HL-60"

/clone_lib="lambda hgc18."

106..1716

/EC_number="3.2.1.45"

/codon_start=1

/product="glucocerebrosidase"

/protein_id="BAA02545.1"

/db_xref="GI:496369"

/translation="MEFSSPSRECPKPLSRVSIMAGSLTGLLLQAVSWASGARPCI
PKSFGYSVWVCVNCATYCDSDPPTFPALGTFSRYESRSGRRMELSMGPIQANHTGT
GLLLTLQPEQKQKVKVGGGAMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIRV
PMASCDFSIRTYTADTDFDLHNFSLPEEDTKLTIPLIHRALQLAQRPVSVLLASPW
TSPTWLKTNAGVNGKSLKGOPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPS
AGLLSGYFQCLGTFPEHQDFIARDLPTLANSTHNVRLMLDDQLLPHWAKVY
LTDPEAKYVHGVIAVHWYLDLAPAKATIGETHRLEFPNTMLFASEACVSKFWEQSVR
LGSWDRGMQYSHSIITNLLYHVVGWTDWNLALNPEGGNVVRNFVDSPLIIVDITKDTF
YKQPMFYHLGHFSKFIPESQVRGLVASQKNDLDAVALMHPDGSVAVVVLNRSKDV
LTIKDPAVGFLETISPQYSIHTYLWRRQ"

106..222

sig_peptide

mat_peptide

/product="glucocerebrosidase"

polyA_site 2259

BASE COUNT 517 a 648 c 598 g 496 t

ORIGIN

Query Match 98.5%; Score 2245.4; DB 9; Length 2259;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AGCTAAGCAGGTACCTGCATCCTTGTGTTTGTGTTAGTGGATCCTCTATCCTTCAGAGAC 60
Db 1 AGCTAAGCAGGTACCTGCATCCTTGTGTTTGTGTTAGTGGATCCTCTATCCTTCAGAGAC 60
QY 61 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGGATGGAGTTTCAAGT 120
Db 61 TCTGGAACCCCTGTGGTCTTCTCTCATCTAATGACCCCTGAGGGGATGGAGTTTCAAGT 120
QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180
Db 121 CCTTCCAGAGAGGAATGTCCCAAGCCCTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180
QY 181 ACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCCCGCCCTGCATCCCT 240
Db 181 ACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCCCGCCCTGCATCCCT 240
QY 241 AAAAGCTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATACGTGACTCCCTTT 300
Db 241 AAAAGCTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATACGTGACTCCCTTT 300
QY 301 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGG 360
Db 301 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGG 360


```
QY 361 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 420
Db 361 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 420
QY 421 CTGACCCCTGCAGCCAGAACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 480
Db 421 CTGACCCCTGCAGCCAGAACAGAAAGTTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 480
QY 481 GATGCTGCTCTCAACATCCTTGCCCTGTCACCCCTGCCCCAAATTTGCTACTTAAA 540
Db 481 GATGCTGCTCTCAACATCCTTGCCCTGTCACCCCTGCCCCAAATTTGCTACTTAAA 540
QY 541 TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 600
Db 541 TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 600
QY 601 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAAC 660
Db 601 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAAC 660
QY 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCCGAGCCCTGCAG 720
Db 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCCGAGCCCTGCAG 720
QY 721 TTGGCCCGAGCGTCCGTTTCACTCCTTGCCAGCCCTGACATACCCACTTGGCTCAAG 780
Db 721 TTGGCCCGAGCGTCCGTTTCACTCCTTGCCAGCCCTGACATACCCACTTGGCTCAAG 780
QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
Db 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
QY 841 CAGACCTGGGCCAGATACCTTTGTGAAGTTCTTGATGCCTATGCTGAGCACAAGTTACAG 900
Db 841 CAGACCTGGGCCAGATACCTTTGTGAAGTTCTTGATGCCTATGCTGAGCACAAGTTACAG 900
QY 901 TTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTC 960
Db 901 TTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTC 960
QY 961 CAGTGCCTGGCTTACCCCTGAACATCAGCAGACTTCAATGCCCCTGACCTAGGTCCT 1020
Db 961 CAGTGCCTGGCTTACCCCTGAACATCAGCAGACTTCAATGCCCCTGACCTAGGTCCT 1020
QY 1021 ACCCTGCCCCAAGTACTCACCACAAATGCGCCCTACTCATGCTGATGACCAACGCTTG 1080
Db 1021 ACCCTGCCCCAAGTACTCACCACAAATGCGCCCTACTCATGCTGATGACCAACGCTTG 1080
QY 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCTAAATATGTTTCA 1140
Db 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCTAAATATGTTTCA 1140
QY 1141 GGCATTGCTGTACATGTTGATACCTGGACTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1200
Db 1141 GGCATTGCTGTACATGTTGATACCTGGACTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1200
QY 1201 ACACACCGCCTGTCCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTCCAAG 1260
Db 1201 ACACACCGCCTGTCCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTCCAAG 1260
QY 1261 TTCTGGGAGCAGAGTGTGCGGCTAGGCTCTGCGGATCGAGGGATGACAGCCACAGC 1320
Db 1261 TTCTGGGAGCAGAGTGTGCGGCTAGGCTCTGCGGATCGAGGGATGACAGCCACAGC 1320
QY 1321 ATCATCACGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGGAACCTTGCCCTGAAC 1380
Db 1321 ATCATCACGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGGAACCTTGCCCTGAAC 1380
QY 1381 CCCGAAGGAGGACCAATTTGGGTGCGTAACTTTGTGACAGATCCCATCATTTGTAGACATC 1440
Db 1381 CCCGAAGGAGGACCAATTTGGGTGCGTAACTTTGTGACAGATCCCATCATTTGTAGACATC 1440
```

```
QY 1441 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGCCACTTCAGCAAGTTC 1500
Db 1441 ACCAAGGACACGTTTACAAACAGCCCATGTTCTACACCTTGGCCACTTCAGCAAGTTC 1500
QY 1501 ATTCTTGAGGGCTCCCAGAGAGTGGGGCTGGTTCAGAGTCAGAAGAACGACCTGGACGCA 1560
Db 1501 ATTCTTGAGGGCTCCCAGAGAGTGGGGCTGGTTCAGAGTCAGAAGAACGACCTGGACGCA 1560
QY 1561 GTGGCACTGATGATCCCGATGGCTCTGCTGTTGTGGTGGTCTAAACCGCTCTCTAAG 1620
Db 1561 GTGGCACTGATGATCCCGATGGCTCTGCTGTTGTGGTGGTCTAAACCGCTCTCTAAG 1620
QY 1621 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAGACAATCTACCTGGC 1680
Db 1621 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAGACAATCTACCTGGC 1680
QY 1681 TACTCCATTACACACCTACCTGTGGCTGCGCCAGTGCAGTGCAGATCTCAAGGAGGCAC 1740
Db 1681 TACTCCATTACACACCTACCTGTGGCTGCGCCAGTGCAGTGCAGATCTCAAGGAGGCAC 1740
QY 1741 TGGGCTCAGCCTGGGCATTAAGGGACAGAGTCACTCACACGCTGTCTGTGACTAAAGA 1800
Db 1741 TGGGCTCAGCCTGGGCATTAAGGGACAGAGTCACTCACACGCTGTCTGTGACTAAAGA 1800
QY 1801 GGGCACAGCAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGTTGGGT 1860
Db 1801 GGGCACAGCAGGGCCAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGTTGGGT 1860
QY 1861 GACTCACTTTCCCTCTAGTGGTG - CCAGGGGCTGGAGGCCCTTAGAAAAGATCAGTA 1919
Db 1861 GACTCACTTTCCCTCTAGTGGTG - CCAGGGGCTGGAGGCCCTTAGAAAAGATCAGTA 1920
QY 1920 AGCCCCAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGTGTGTGCTGCTT 1979
Db 1921 AGCCCCAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGTGTGTGCTGCTT 1980
QY 1980 TGGAAACTGGGCTGGTCCAGGCTTAGGGTGAAGTCACTGTCCGTACAAACACAAAGATC 2039
Db 1981 TGGAAACTGGGCTGGTCCAGGCTTAGGGTGAAGTCACTGTCCGTACAAACACAAAGATC 2040
QY 2040 AGGCTGAGGGTAAGGAAAGAGAGACTAGGAAAGTGGGCCCAAACTGGAGACTGTT 2099
Db 2041 AGGCTGAGGGTAAGGAAAGAGAGACTAGGAAAGTGGGCCCAAACTGGAGACTGTT 2100
QY 2100 TGTCTTTTCTGGAGATGCAAACTGGGCTGGGCTGGAGCAGTGTGATGATGAGGCGGA 2159
Db 2101 TGTCTTTTCTGGAGATGCAAACTGGGCTGGGCTGGAGCAGTGTGATGATGAGGCGGA 2160
QY 2160 AGCCTTAAAGCAGCAGCGGTGTGCCAGGACCCAGCAGATGATTCCTATGGCAGCCAG 2219
Db 2161 AGCCTTAAAGCAGCAGCGGTGTGCCAGGACCCAGCAGATGATTCCTATGGCAGCCAG 2220
QY 2220 GAAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGAGCCC 2258
Db 2221 GAAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGAGCCC 2259
```

```
RESULT 3
HUMCBPRC HUMCBPRC 2275 bp mRNA linear PRI 27-APR-1993
LOCUS Human glucocerebrosidase mRNA, complete cds.
DEFINITION M19285
ACCESSION M19285
VERSION M19285.1 GI:183027
KEYWORDS glucocerebrosidase.
SOURCE Human erythroid cell line K562, cDNA to mRNA, clone p7-2-1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2275)
AUTHORS Horowitz,M.
JOURNAL Unpublished (1988)
REFERENCE 2 (bases 1 to 140)
AUTHORS Reiner,O., Wigderson,M. and Horowitz,M.
```


TITLE Structural analysis of the human glucocerebrosidase genes
 JOURNAL DNA 7 (2), 107-116 (1988)
 MEDLINE 88195776
 PUBMED 3359914
 COMMENT Draft entry and computer-readable sequence for [2],[1] kindly provided by M.Horowitz, 13-MAY-1988.
 An alternative initiation codon is locate at positions 183-185 and alternative poly-adenylation signals at positions 2250-2255 (major) and 1774-1779 (minor).
 FEATURES
 source Location/Qualifiers
 1..2275
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 <1..2275
 /product="GCB mRNA"
 123..1733
 /note="glucocerebrosidase"
 /codon_start=1
 /protein_id="AAA35880.1"
 /db_xref="GI:183028"
 /translation="MEFSSPSREECPKPLSRVSIMAGSLTGLLLQAVSWASGARPCI
 PKSGYSSVVCVNCATYCDSEDPPTPALGTFSRYESTRSRRMELSGPIQANHTGT
 GLLLTLOPEOKFKQKVGFGGAMTDAALNILALSPRAQNLKLSYFSEEGIGYNIIRV
 PMASCTFSIRTYTADTPDDQLHNFSLPEEDTKLKIPLIHRALQIAORPVSLIASPW
 TSPDWLKTNGAVNGKSLGQPGDIYHQTWARYFVKLDAYAEHKLQFWAVTAENEPS
 AGLLSGYPFOCLGFTPEHQRFDIARDLPTLANSTHNVRLMLDDQRLLLPHWAKVY
 LTDPEAAKYVHGIADVHVLDFLAPAKATLGETHRLFNTMLFASEACVSGSKFWEQSVR
 LGSWDRGMQYSHSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIIVDITKDF
 YKQPMFYHLGHFISKFIPESQVRGLVASQKNDLDAVALMHPDGSVVVVLNRSKDV
 LTIKDPVAVGELETISPGYSIHTYLWHRQ"
 BASE COUNT 523 a 652 c 596 g 503 t 1 others
 ORIGIN 32 bp upstream of KpnI site; chromosome 1q21-31.
 Query Match 98.2%; Score 2238; DB 9; Length 2275;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2252; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 QY 1 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTAGTGATCCTCTATCCTTCAGAGAC 60
 DB 18 AGCTAAGGCAGGTACCTGCATCCTTGTGTTTGTAGTGATCCTCTATCCTTCAGAGAC 77
 QY 61 TCTGAACCCCTGTGCTCTTCTTTCATCTAATGACCTGAGGGGATGGAGTTTCAAGT 120
 DB 78 TCTGAACCCCTGTGCTCTTCTTTCATCTAATGACCTGAGGGGATGGAGTTTCAAGT 137
 QY 121 CCTTCCAGAGAGGAATGCCAAGCCTTTGAGTAGGTAAGCATCATATGGCTGGCAGCCTC 180
 DB 138 CCTTCCAGAGAGGAATGCCAAGCCTTTGAGTAGGTAAGCATCATATGGCTGGCAGCCTC 197
 QY 181 ACAGGATGCTTCTACTTCAGGCAGTGTCTGTTGGGCATCAGGTGCCCGCCCTGCATCCCT 240
 DB 198 ACAGGTTTGTCTTACTTCAGGCAGTGTCTGTTGGGCATCAGGTGCCCGCCCTGCATCCCT 257
 QY 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATFACTGTGACTCCTTT 300
 DB 258 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATFACTGTGACTCCTTT 317
 QY 301 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGCGAGTGGG 360
 DB 318 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGCGAGTGGG 377
 QY 361 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 420
 DB 378 CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTA 437
 QY 421 CTGACCCCTGCAGCCAGAACAGAGTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 480
 DB 438 CTGACCCCTGCAGCCAGAACAGAGTCCAGAAAGTGAAGGGATTGGAGGGGCCATGACA 497
 QY 481 GATGCTGCTGCTCAACATCCTTGCCCTGTCTACCCCTGTCTACCCCAAAATTTGCTACTTAA 540
 DB 498 GATGCTGCTGCTCAACATCCTTGCCCTGTCTACCCCTGTCTACCCCAAAATTTGCTACTTAA 557

QY 541 TCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCGGGTACCCATGGCCAGCTGT 600
 DB 558 TCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCGGGTACCCATGGCCAGCTGT 617
 QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAAC 660
 DB 618 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAAC 677
 QY 661 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
 DB 678 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 737
 QY 721 TTGGCCCGAGCGTCCCGTTTCACTCCTTGCAGCCCTGAGCATCACCCACTTGGCTCAAG 780
 DB 738 TTGGCCCGAGCGTCCCGTTTCACTCCTTGCAGCCCTGAGCATCACCCACTTGGCTCAAG 797
 QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
 DB 798 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 857
 QY 841 CAGACCTGGCCAGATACCTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 900
 DB 858 CAGACCTGGCCAGATACCTTGTGAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 917
 QY 901 TTCTGGGAGTGACAGCTGAAATGAGCCTTCTGCTGGGTGTTGAGTGGATACCCCTTC 960
 DB 918 TTCTGGGAGTGACAGCTGAAATGAGCCTTCTGCTGGGTGTTGAGTGGATACCCCTTC 977
 QY 961 CAGTGCCTGGGCTTCACCCCTGAACATCAGCGAGACTTTCATTTGCCCGTACCTAGGTCCT 1020
 DB 978 CAGTGCCTGGGCTTCACCCCTGAACATCAGCGAGACTTTCATTTGCCCGTACCTAGGTCCT 1037
 QY 1021 ACCCTCGCCAAACAGTACTACCAACAATGTCGCCCTACTCATGCTGGATGACCAACGCTTG 1080
 DB 1038 ACCCTCGCCAAACAGTACTACCAACAATGTCGCCCTACTCATGCTGGATGACCAACGCTTG 1097
 QY 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAAAGCAGCTAAATATGTTTCA 1140
 DB 1098 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAAAGCAGCTAAATATGTTTCA 1157
 QY 1141 GGCATTGCTGTACATTGGTACCTGGACTTCTTGGCTCCAGCCAAAGCCACCTTAGGGGAG 1200
 DB 1158 GGCATTGCTGTACATTGGTACCTGGACTTCTTGGCTCCAGCCAAAGCCACCTTAGGGGAG 1217
 QY 1201 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCCCTCAGAGGCCCTGTGTGGGCTCCAAG 1260
 DB 1218 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCCCTCAGAGGCCCTGTGTGGGCTCCAAG 1277
 QY 1261 TTCTGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGC 1320
 DB 1278 TTCTGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGC 1337
 QY 1321 ATCATACGAAACCTCCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGCCCTGAAC 1380
 DB 1338 ATCATACGAAACCTCCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGCCCTGAAC 1397
 QY 1381 CCCGAAGGAGGACCCCAATTGGGTGCGTAACTTTGTCACAGTCCCATCATTTGTAGACATC 1440
 DB 1398 CCCGAAGGAGGACCCCAATTGGGTGCGTAACTTTGTCACAGTCCCATCATTTGTAGACATC 1457
 QY 1441 ACCAAGGACACGTTTACAAAACAGCCCATGTTTACACACCTTTGGCCACTTCAGCAAGTTC 1500
 DB 1458 ACCAAGGACACGTTTACAAAACAGCCCATGTTTACACACCTTTGGCCACTTCAGCAAGTTC 1517
 QY 1501 ATTCTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAAGAACGACCTGGAGCGCA 1560
 DB 1518 ATTCTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAAGAACGACCTGGAGCGCA 1577
 QY 1561 GTGCACTGATGATCCCGATGGCTCTGCTGTTGTGTGCTGCTAAACCGCTCCTCTAAG 1620
 DB 1578 GTGCACTGATGATCCCGATGGCTCTGCTGTTGTGTGCTGCTAAACCGCTCCTCTAAG 1637
 QY 1621 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTTGGAGACAATCTCACCTGGC 1680

```
|||||
Db 1638 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGCTTCCTGGAGACAATCTCACCTGSC 1697
QY 1681 TACTCCATTACACCTACCTGTGGCGTCGCCAGTGATGGAGCAGATACCTCAAGGAGGCAC 1740
Db 1698 TACTCCATTACACCTACCTGTGGCATCGCCAGTGATGGAGCAGATACCTCAAGGAGGCAC 1757
QY 1741 TGGGCTCAGCTGGGCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGA 1800
Db 1758 TGGGCTCAGCTGGGCATTAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGA 1817
QY 1801 GGGCACAGCAGGSCCAGTGTGAGCTTACAGCGACGTAAAGCCAGGGGCAATGGTTGGGT 1860
Db 1818 GGGCACAGCAGGSCCAGTGTGAGCTTACAGCGACGTAAAGCCAGGGGCAATGGTTGGGT 1877
QY 1861 GACTCACCTTCCCTCTAGGTGCTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAA 1920
Db 1878 GACTCACCTTCCCTCTAGGTGCTGCCAGGGGCTGGAGGCCCTAGAAAAAGATCAGTAA 1937
QY 1921 GCCCCAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTGCTTT 1980
Db 1938 GCCCCAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTGCTTT 1997
QY 1981 GGAACTGGGCTGGGTCAGGCTAGGCTGAGCTCACTGTCCGTACAAACACAAGATCA 2040
Db 1998 GGAACT - NGCCTGGGTCCAGGCTAGGCTGAGCTCACTGTCCGTACAAACACAAGATCA 2056
QY 2041 GGGCTGAGGTAAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCCAAAAGCTGGAGACTGTTT 2100
Db 2057 GGGCTGAGGTAAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCCAAAAGCTGGAGACTGTTT 2116
QY 2101 GTCTTTCTGAGATGCAGAACTGGGCGCTGGAGCAGCAGTGTCAGCATCAGGGCGGAA 2160
Db 2117 GTCTTTCTGAGATGCAGAACTGGGCGCTGGAGCAGCAGTGTCAGCATCAGGGCGGAA 2176
QY 2161 GCCTTAAAGCAGCAGCGGGTGTGCCAGGCCAGCCAGATGATTCCTATGGCAGCCAGG 2220
Db 2177 GCCTTAAAGCAGCAGCGGGTGTGCCAGGCCAGCCAGATGATTCCTATGGCAGCCAGG 2236
QY 2221 AAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGGGCCCA 2259
Db 2237 AAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGGGCCCA 2275
```

```
RESULT 4
HUMGB HUMGB 2587 bp mRNA linear PRI 27-APR-1993
LOCUS Human glucocerebrosidase mRNA, complete cds.
DEFINITION M16328 M11080
ACCESSION M16328.1 GI:183007
VERSION 1
KEYWORDS glucocerebrosidase.
SOURCE Human fibroblast cell line WI-38, CDNA to mRNA, clones G5A-1Y and ID9-bb.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2587)
AUTHORS Sorge,J., West,C., Westwood,B. and Beutler,E.
TITLE Molecular cloning and nucleotide sequence of human
glucocerebrosidase cDNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82 (21), 7289-7293 (1985)
MEDLINE 86042651
PUBMED 3864160
REFERENCE 2 (bases 1 to 2587)
AUTHORS Sorge,J., Kuhl,W., West,C. and Beutler,E.
TITLE Gaucher disease: retrovirus-mediated correction of the enzymatic
defect in cultured cells
JOURNAL Cold Spring Harb. Symp. Quant. Biol. 51 Pt 2, 1041-1046 (1986)
MEDLINE 87217081
PUBMED 3472750
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by J.Sorge, 23-JAN-1986.
Mutations in the human glucocerebrosidase gene cause Gaucher
```

disease. Affected individuals cannot adequately catabolize glucocerebroside. Authors have not been able to establish a definitive 5' end for the glucocerebrosidase mRNA.

FEATURES
source Location/Qualifiers
1..2587 /organism="Homo sapiens"
/db_xref="taxon:9606"
mRNA <1..2587
/product="gcb mRNA"
496..2046 /note="glucocerebrosidase precursor (5' end put.); putative"

/codon_start=1
/protein_id="AAA35873.1"
/db_xref="GI:183008"
/translation="MAGSLTGLLLQAVSWAGARPCIPKSFYSSVVCVNCATYCDSDPPTFPALGTFSTRYESTSRGRMELSMGP IQANHTGTGLLTLOPEQKFKVGFGGAMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDIFSIRTYTADTPDDFQHNFSLPEEDTKLKIPLIHRAIQLAQPVPVSLASPTSPWLKTNAGAVNGKSLKGQPGDIHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPFOCLGFTPEHQRFIARDLGPPTLANSTHNVRLLMDQRLLLPHWAKVVLTPDEAAKYVHGIADVWYLD FLAPAKATLGETHRLFPNTMLFASEACVGSKEQSVRLGSDRGMQYSHSIITNLLYHVVGWTDNWLALNPEGGPVWVNFVDSPIIVDITKDFYKQPMFYHLGHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAVGFLETISPGLYSIHTYLWHRQ"
sig_peptide 496..552 /note="glucocerebrosidase signal peptide (5' end put.); putative"

mat_peptide 553..2043 /product="glucocerebrosidase"

BASE COUNT 635 a 710 c 654 g 587 t 1 others
ORIGIN 8 bp upstream of StuI site.

Query Match 97.4%; Score 2220; DB 9; Length 2587;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2234; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
QY 18 GCATCCTTGTGTTTGTAGTGGATCCTCTATCCTCAGAGACTCTGGAACCCCTGTGGT 77
Db 348 GCATCCTTGTGTTTGTAGTGGATCCTCTATCCTCAGAGACTCTGGAACCCCTGTGGT 407
QY 78 CTCTCTTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGTCCCTCCAGAGAGGAATG 137
Db 408 CTCTCTTCATCTAATGACCCCTGAGGGATGGAGTTTCAAGTCCCTCCAGAGAGGAATG 467
QY 138 TCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCCTCACAGGATTCGTTCTACT 197
Db 468 TCCCAAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCCTCACAGGTTTGTCTACT 527
QY 198 TCAGGCAGTGTGCTGGGCATCAGGTGCCGCCCTTCATCCCTTAAAGCTTCGGCTACAG 257
Db 528 TCAGGCAGTGTGCTGGGCATCAGGTGCCGCCCTTCATCCCTTAAAGCTTCGGCTACAG 587
QY 258 CTCGGTGGTGTGTCTGCAATGCCACATACCTGTGACTCCTTTGACCCCCGACCTTCC 317
Db 588 CTCGGTGGTGTGTCTGCAATGCCACATACCTGTGACTCCTTTGACCCCCGACCTTCC 647
QY 318 TGCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGCGACGGATGGAGCTGAG 377
Db 648 TGCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGTGGCGACGGATGGAGCTGAG 707
QY 378 TATGGGGCCCATCCAGGCTAATCAGGGGACAGCCGCTGCTACTGACCCCTGCAGCCAGA 437
Db 708 TATGGGGCCCATCCAGGCTAATCAGGGGACAGCCGCTGCTACTGACCCCTGCAGCCAGA 767
QY 438 ACAGAACTCCAGAAAGTGAAGGATTTGGAGGGGCCATGACAGATGCTGCTGCTCTCAA 497
Db 768 ACAGAACTCCAGAAAGTGAAGGATTTGGAGGGGCCATGACAGATGCTGCTGCTCTCAA 827
QY 498 CATCCTTGCCTGTGCACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGA 557
Db 828 CATCCTTGCCTGTGCACCCCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGA 887
```


| | | | |
|----|------|--|------|
| QY | 558 | AGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCTCCATCCGCAC | 617 |
| Db | 888 | AGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCTCCATCCGCAC | 947 |
| QY | 618 | CTACACCTATGCACACACCCCTGATGATTTCCAGTTGCACAACTTCAGCCTCCCAGAGGA | 677 |
| Db | 948 | CTACACCTATGCACACACCCCTGATGATTTCCAGTTGCACAACTTCAGCCTCCCAGAGGA | 1007 |
| QY | 678 | AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCACGCGTCCCGT | 737 |
| Db | 1008 | AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCACGCGTCCCGT | 1067 |
| QY | 738 | TTCACTCCTTGCCAGCCCCCTGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAA | 797 |
| Db | 1068 | TTCACTCCTTGCCAGCCCCCTGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAA | 1127 |
| QY | 798 | TGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACCAGACCTGGGCCAGATA | 857 |
| Db | 1128 | TGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACCAGACCTGGGCCAGATA | 1187 |
| QY | 858 | CTTTGTGAAGTTCCTGGATGCCATGCTGAGCAACAAGTTACAGTTCTGGGCAGTGACAGC | 917 |
| Db | 1188 | CTTTGTGAAGTTCCTGGATGCCATGCTGAGCAACAAGTTACAGTTCTGGGCAGTGACAGC | 1247 |
| QY | 918 | TGAAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC | 977 |
| Db | 1248 | TGAAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC | 1307 |
| QY | 978 | CCCTGAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCTTACCCTCGCCAACAGTAC | 1037 |
| Db | 1308 | CCCTGAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCTTACCCTCGCCAACAGTAC | 1367 |
| QY | 1038 | TCACCACAATGTCCGCCCTACTCATGCTGGATGACCAACGCTTGCTGCTGCCCCACATGGGC | 1097 |
| Db | 1368 | TCACCACAATGTCCGCCCTACTCATGCTGGATGACCAACGCTTGCTGCTGCCCCACATGGGC | 1427 |
| QY | 1098 | AAAGGTGGTACTGACAGACCCAGAACGACGCTAAATATGTTCAATGGCATTGTGTACATTG | 1157 |
| Db | 1428 | AAAGGTGGTACTGACAGACCCAGAACGACGCTAAATATGTTCAATGGCATTGTGTACATTG | 1487 |
| QY | 1158 | GTACCTGGACTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCGCCTGTCCCC | 1217 |
| Db | 1488 | GTACCTGGACTTCTGGCTCCAGCCAAAGCCACCCCTAGGGGAGACACACCGCCTGTCCCC | 1547 |
| QY | 1218 | CAACACCATGCTCTTTGCCCTCAGAGGCCTGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGT | 1277 |
| Db | 1548 | CAACACCATGCTCTTTGCCCTCAGAGGCCTGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGT | 1607 |
| QY | 1278 | CGGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGCATCATCAGAACCTCCT | 1337 |
| Db | 1608 | CGGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGCATCATCAGAACCTCCT | 1667 |
| QY | 1338 | GTACCATGTGGTCGGCTGGACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCAA | 1397 |
| Db | 1668 | GTACCATGTGGTCGGCTGGACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCAA | 1727 |
| QY | 1398 | TTGGGTGGTAACTTTGTGACAGTCCCATCATTTAGACATCACCAAGGACACGTTTATA | 1457 |
| Db | 1728 | TTGGGTGGTAACTTTGTGACAGTCCCATCATTTAGACATCACCAAGGACACGTTTATA | 1787 |
| QY | 1458 | CAAACAGCCCCATGTTCTACCACCTTGCCCACTTCAGCAAGTTCAATTCCTGAGGGCTCCCA | 1517 |
| Db | 1788 | CAAACAGCCCCATGTTCTACCACCTTGCCCACTTCAGCAAGTTCAATTCCTGAGGGCTCCCA | 1847 |
| QY | 1518 | GAGAGTGGGGCTGGTTGCCAGTCAGAAGAACGACCTGGACGCGAGTGGCACTGATGCATCC | 1577 |
| Db | 1848 | GAGAGTGGGGCTGGTTGCCAGTCAGAAGAACGACCTGGACGCGAGTGGCACTGATGCATCC | 1907 |
| QY | 1578 | CGATGGCTCTGCTGTTGTGGTCTGCTAAACCGCTCCTCTAAGGATGTGCCTCTTACCAT | 1637 |
| Db | 1908 | CGATGGCTCTGCTGTTGTGGTCTGCTAAACCGCTCCTCTAAGGATGTGCCTCTTACCAT | 1967 |

| | | | |
|----|------|--|------|
| QY | 1638 | CAAGGATCCTGCTGTGGGCTTCCTGGAGACAAATCTCACCTGGCTACTCCATTTCACACCTA | 1697 |
| Db | 1968 | CAAGGATCCTGCTGTGGGCTTCCTGGAGACAAATCTCACCTGGCTACTCCATTTCACACCTA | 2027 |
| QY | 1698 | CCTGTGGCGTCGCCAGTGATGGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGGGCA | 1757 |
| Db | 2028 | CCTGTGGCATCGCCAGTGATGGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGGGCA | 2087 |
| QY | 1758 | TAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGCAGGGCCAG | 1817 |
| Db | 2088 | TAAAGGGACAGAGTCAGCTCACACGCTGTCTGTGACTAAAGAGGGCACAGCAGGGCCAG | 2147 |
| QY | 1818 | TGTGAGCTTACAGCGACGTTAAGCCCAAGGGCAATGGTTTGGGTGACTCACTTTCCTCTCT | 1877 |
| Db | 2148 | TGTGAGCTTACAGCGACGTTAAGCCCAAGGGCAATGGTTTGGGTGACTCACTTTCCTCTCT | 2207 |
| QY | 1878 | AGGTGTCGCCAGGGGCTGGAGGCCCTAGAAAAAGATCACTAAGCCCCCAGTGTCCCCCA | 1937 |
| Db | 2208 | AGGTGTCGCCAGGGCTGGAGGCCCTAGAAAAAGATCACTAAGCCCCCAGTGTCCCCCA | 2267 |
| QY | 1938 | GCCCCCATGCTTATGTGAACATGCGCTGTGCTGCTTGGAACTGGGCGCTGGGT | 1997 |
| Db | 2268 | GCCCCCATGCTTATGTGAACATGCGCTGTGCTGCTTGGAACT - NGCCTGGGT | 2326 |
| QY | 1998 | CCAGGCCTAGGCTGAGCTCACTGTCCGTACAAACACAAGATCAGGGCTAGGGTAAGGAA | 2057 |
| Db | 2327 | CCAGGCCTAGGCTGAGCTCACTGTCCGTACAAACACAAGATCAGGGCTAGGGTAAGGAA | 2386 |
| QY | 2058 | AAGAAGAGACTAGGAAAGCTGGGCCCCAAAACCTGGAGACTGTTGTCTTCTTGAGATGC | 2117 |
| Db | 2387 | AAGAAGAGACTAGGAAAGCTGGGCCCCAAAACCTGGAGACTGTTGTCTTCTTAGAGATGC | 2446 |
| QY | 2118 | AGAACTGGGCCCCGTGGAGCAGCAGTGTCAGCATCAGGGCGGAAGCCTTAAAGCAGCAGCG | 2177 |
| Db | 2447 | AGAACTGGGCCCCGTGGAGCAGCAGTGTCAGCATCAGGGCGGAAGCCTTAAAGCAGCAGCG | 2506 |
| QY | 2178 | GGTGTGCCAGGCCACCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTCTTA | 2237 |
| Db | 2507 | GGTGTGCCAGGCCACCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTCTTA | 2566 |
| QY | 2238 | AAGGAGAAATGTTTGAGCCC | 2258 |
| Db | 2567 | AAGGAGAAATGTTTGAGCCC | 2587 |

| | |
|------------|--|
| RESULT 5 | |
| HUMGCBL | |
| LOCUS | HUMGCBL |
| DEFINITION | Human lysosomal glucocerebrosidase mRNA, complete cds. |
| ACCESSION | K02920 |
| VERSION | K02920.1 GI:183021 |
| KEYWORDS | glucocerebrosidase. |
| SOURCE | Human hepatoma, cDNA to mRNA, clone pGC1. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | 1 (bases 1 to 1792) Tsuiji, S., Choudary, P.V., Martin, B.M., Winfield, S., Barranger, J.A. and Ginns, E.I. |
| TITLE | Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase |
| JOURNAL | J. Biol. Chem. 261 (1), 50-53 (1986) |
| MEDLINE | 86085859 |
| PUBMED | 3001061 |

| COMMENT | REMARKS |
|--|----------------|
| <p>A preprint for [1] was kindly provided by E.I.Ginnis, 01-AUG-1985</p> <p>Glucocerebrosidase is a membrane bound lysosomal enzyme, which catalyzes the hydrolysis of glucocerebrosidase. Allelic mutations in the gene coding for glucocerebrosidase cause the alteration of normal synthesis and/or processing, resulting in sphingolipidosis. The precise nature of the genetic changes resulting in the three phenotypes of this disease have not yet been determined. The glucocerebrosidase sequence reported by [1] differs by one AA from human placental glucocerebrosidase.</p> | <p>5001001</p> |

FEATURES
source
Location/Qualifiers
1..1792
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>1792
/product="lgc mRNA"
46..1596
CDS
/note="lysosomal glucocerebrosidase precursor, EC 3.2.1.45"
/codon_start=1
/protein_id="AAA35877.1"
/db_xref="GI:183022"
/translation="MAGSLTGLLLQLQAVSWASGARPICPKSFGYSVVVCNATYCDSDPPTFPALGTFSTRSRRRMELSMGPIQANHTGTGLLLTLQPEQKQKVGFGGAMTDAALNIALSPPAQNLLKSYFSEEGIGNIIRVPMASCDFSIRTYADTPDDFOLHNFSLPEEDTKLIPLIHRAQLAQRPSVLSLASPWTSPWLKTNAGVNGKSLKGOPGDIYHQTWARYFVKFLDAYAEHLQFWAVTAENEPSAGLLSGYPFOCLGFTPEHORDLIADLGPPTLANSTHNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGLAVHWYLDFLAPAKATLGETHRLFPNTMLFASACVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDNLANLNPGEPPNVRNFVDSPIIIVDITKDFYKQPMFYHLGHSKFIPEGSQRVGLVASQKNDLDAVALMHPDGSVAVVVVNLNRSSKDVPILTIKDPVAVGFLETISPYSIHTYLWRRQ"
sig_peptide
46..102
/note="lysosomal glucocerebrosidase signal peptide"
mat_peptide
103..1593
/product="lysosomal glucocerebrosidase"
BASE COUNT 406 a 528 c 466 g 392 t
ORIGIN 311 bp upstream of PstI site; chromosome 1q21.
Query Match 78.4%; Score 1785.6; DB 9; Length 1792;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1788; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 121 CCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180
Db 1 CCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGGTAAGCATCATGGCTGGCAGCCTC 60
QY 181 ACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGGCATCAGGTGCCCGCCCTGCATCCCT 240
Db 61 ACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGGCATCAGGTGCCCGCCCTGCATCCCT 120
QY 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
Db 121 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY 301 GACCCCGGACCTTTCTGCTGCTTGGTACCTTCAGCCGCTATGAGAGTACACGCGAGTGGG 360
Db 181 GACCCCGGACCTTTCTGCTGCTTGGTACCTTCAGCCGCTATGAGAGTACACGCGAGTGGG 240
QY 361 CGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACGGGCGACAGGCTGTCTA 420
Db 241 CGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACGGGCGACAGGCTGTCTA 300
QY 421 CTGACCCCTGCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTTGGAGGGGCCATGACA 480
Db 301 CTGACCCCTGCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTTGGAGGGGCCATGACA 360
QY 481 GATGCTGCTGCTCTCAACATCCTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAA 540
Db 361 GATGCTGCTGCTCTCAACATCCTTGGCCCTGTACCCCTGCCCCAAATTTGCTACTTAA 420
QY 541 TCCTACTTCTCTGAAGAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 600
Db 421 TCCTACTTCTCTGAAGAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT 480
QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGATTTCCAGTTGCACAAC 660
Db 481 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGATTTCCAGTTGCACAAC 540
QY 661 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGTATTCACCGACCCCTGCAG 720
Db 541 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGTATTCACCGACCCCTGCAG 600

QY 721 TTGGCCAGCGTCCCGTTTCACTCCTTCCAGCCCTGGACATCACCACCTTGGCTCAAG 780
Db 601 TTGGCCAGCGTCCCGTTTCACTCCTTCCAGCCCTGGACATCACCACCTTGGCTCAAG 660
QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
Db 661 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 720
QY 841 CAGACCTGGGCCAGATACCTTTGTAAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 900
Db 721 CAGACCTGGGCCAGATACCTTTGTAAAGTTCTTGGATGCCCTATGCTGAGCACAAGTTACAG 780
QY 901 TTCTGGGCAGTGACAGCTGAAAAATGAGCCTTCTTGGTGGCTGTGTGAGTGATACCCCTTC 960
Db 781 TTCTGGGCAGTGACAGCTGAAAAATGAGCCTTCTTGGTGGCTGTGTGAGTGATACCCCTTC 840
QY 961 CAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCAATGCCCGTGACCTAGGTCCT 1020
Db 841 CAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTAAATGCCCGTGACCTAGGTCCT 900
QY 1021 ACCCTGCCCAACAGTACTCACCACAATGTCCGCCTACTCATGCTGGATGACCAACGCTTG 1080
Db 901 ACCCTGCCCAACAGTACTCACCACAATGTCCGCCTACTCATGCTGGATGACCAACGCTTG 960
QY 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAAATATGTTTCAT 1140
Db 961 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAAATATGTTTCAT 1020
QY 1141 GGCATTGCTGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1200
Db 1021 GGCATTGCTGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTAGGGGAG 1080
QY 1201 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTCCAAG 1260
Db 1081 ACACACCGCCTGTTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTCCAAG 1140
QY 1261 TTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGATCGAGGGATGCAGTACAGCCACAGC 1320
Db 1141 TTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGATCGAGGGATGCAGTACAGCCACAGC 1200
QY 1321 ATCATCACGAACCTCCTGTACCATGTGGTGGCTGGAGCAGCTGGAACCTTGGCCTGAAC 1380
Db 1201 ATCATCACGAACCTCCTGTACCATGTGGTGGCTGGAGCAGCTGGAACCTTGGCCTGAAC 1260
QY 1381 CCGAAGGAGGACCCCAATTTGGTGCCTAACTTTGTCAGACAGTCCCACATCATTTGAGACATC 1440
Db 1261 CCGAAGGAGGACCCCAATTTGGTGCCTAACTTTGTCAGACAGTCCCACATCATTTGAGACATC 1320
QY 1441 ACCAAGGACACGTTTACAAACAGACCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTC 1500
Db 1321 ACCAAGGACACGTTTACAAACAGACCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTC 1380
QY 1501 ATTCTGAGGGCTCCCAGAGAGTGGGGCTGGTTGCCAGTCAGAGAAGACGACCTGGACGCA 1560
Db 1381 ATTCTGAGGGCTCCCAGAGAGTGGGGCTGGTTGCCAGTCAGAGAAGACGACCTGGACGCA 1440
QY 1561 GTGGCAGTGTATGATCCCGATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
Db 1441 GTGGCAGTGTATGATCCCGATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 1621 GATGCTGCTTACCATCAAGGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
Db 1501 GATGCTGCTTACCATCAAGGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
QY 1681 TACTCCATTCACACCTACCTGTGGCTGCCAGTGTGATGGAGAGATCTCAAGGAGGCAC 1740
Db 1561 TACTCCATTCACACCTACCTGTGGCTGCCAGTGTGATGGAGAGATCTCAAGGAGGCAC 1620
QY 1741 TGGGCTCAGCCTGGGCATTAAAGGACAGAGTCAAGTCAACAGCTGTGTGTGTGTGTGTGT 1800
Db 1621 TGGGCTCAGCCTGGGCATTAAAGGACAGAGTCAAGTCAACAGCTGTGTGTGTGTGTGTGT 1680
QY 1801 GGGCACAGCAGGGGCCAGTGTGAGCTTACAGCGCAGCTAAGCCAGGGCAATGTTTGGGT 1860

Db 1681 GGGCACAGAGGGCCAGTGTAGCTTACAGCGACGTAAAGCCAGGGCCATGGTTTGGGT 1740

QY 1861 GACTCACTTCCCTCTAGGTGGTCCAGGGGGCTGGAGGCCCTAGAAAAAG 1912

Db 1741 GACTCACTTCCCTCTAGGTGGTCCAGGGGGCTGGAGGCCCTAGAAAAAG 1792

RESULT 6

LOCUS I09351 1661 bp DNA linear PAT 02-DEC-1994

DEFINITION Sequence 1 from Patent WO 8905850.

ACCESSION I09351

VERSION I09351.1 GI:587940

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1661)

AUTHORS Ginds, E.I., Martin, B., Maysack, K.A., Eliason, W.K. and LaMarca, M.E.

TITLE CLONED DNA FOR SYNTHESIZING UNIQUE GLUCOCEREBROSIDASE

JOURNAL Patent: WO 8905850-A 1 29-JUN-1989;

FEATURES

source 1. .1661

BASE COUNT 374 a 496 c 422 g 369 t

ORIGIN

Query Match 72.0%; Score 1640.2; DB 6; Length 1661;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1648; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 108 GGAGTTTCAAGTCCCTCCAGAGAGGAATGTCCCAAGCCCTTTGAGTAGGGTAAGCATCAT 167

Db 1 GGAGTTTCAAGTCCCTCCAGAGAGGAATGTCCCAAGCCCTTTGAGTAGGGTAAGCATCAT 60

QY 168 GGCTGGCAGCCCTCACAGGATTTGCTTCTACTTTCAGGCAGTGTCTGTTGGGCATCAGGTGCCG 227

Db 61 GGCTGGCAGCCCTCACAGGATTTGCTTCTACTTTCAGGCAGTGTCTGTTGGGCATCAGGTGCCG 120

QY 228 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATA 287

Db 121 CCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATA 180

QY 288 CTGTGACTCTCTTGACCCCGGACCTTCTCTGCCCCCTTGGTACCTTTCAGCCCGCTATGAGAG 347

Db 181 CTGTGACTCTCTTGACCCCGGACCTTCTCTGCCCCCTTGGTACCTTTCAGCCCGCTATGAGAG 240

QY 348 TACAGCAGTGGGCGACGGATGGAGTGTAGTATGGGGCCCATCCAGGCTAATCACAGGG 407

Db 241 TACAGCAGTGGGCGACGGATGGAGTGTAGTATGGGGCCCATCCAGGCTAATCACAGGG 300

QY 408 CACAGGCCCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAGAAAGTGAAGGATTTGG 467

Db 301 CACAGGCCCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAGAAAGTGAAGGATTTGG 360

QY 468 AGGGCCCATGACAGATGCTGCTCTCAACATCCTTGCCTGTCAACCCCTGCCCCAAAA 527

Db 361 AGGGCCCATGACAGATGCTGCTCTCAACATCCTTGCCTGTCAACCCCTGCCCCAAAA 420

QY 528 TTTGCTACTTAATCGTACTTCTCTGAAGAGGAATCGGATATACATCATCCGGGTACC 587

Db 421 TTTGCTACTTAATCGTACTTCTCTGAAGAGGAATCGGATATACATCATCCGGGTACC 480

QY 588 CATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATTTT 647

Db 481 CATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATTTT 540

QY 648 CCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 707

Db 541 CCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA 600

QY 708 CCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTCTTGGCCAGCCCTGGACATCACC 767

Db 601 CCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTTGGCAGCCCTGGACATCACC 660

QY 768 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCCG 827

Db 661 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCCCG 720

QY 828 AGACATCTACCAACAGACTGGGCCAGATACTTTGTGAAGTTCTTGATGCTATGCTGA 887

Db 721 AGACATCTACCAACAGACTGGGCCAGATACTTTGTGAAGTTCTTGATGCTATGCTGA 780

QY 888 GCACAAAGTTACAGTTCTTGGSCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTGAG 947

Db 781 GCACAAAGTTACAGTTCTTGGSCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTGAG 840

QY 948 TGGATACCCCTTCCAGTGGCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCG 1007

Db 841 TGGATACCCCTTCCAGTGGCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCCCG 900

QY 1008 TGACCTAGGTCTTACCTCGCCCAACAGTACTACCAATGTCCGCTTCTGCTGCTGTTGAG 1067

Db 901 TGACCTAGGTCTTACCTCGCCCAACAGTACTACCAATGTCCGCTTCTGCTGCTGTTGAG 960

QY 1068 TGACCAACGCTTGTCTGCTGCCCACTGGSCAAAGGTGGTACTGACAGACCCAGAACGAGC 1127

Db 961 TGACCAACGCTTGTCTGCTGCCCACTGGSCAAAGGTGGTACTGACAGACCCAGAACGAGC 1020

QY 1128 TAAATATGTTTCAATGGCTGTGTACATTTGGTACCTTGGTCTTCTGGTCCAGCCAAAGC 1187

Db 1021 TAAATATGTTTCAATGGCTGTGTACATTTGGTACCTTGGTCTTCTGGTCCAGCCAAAGC 1080

QY 1188 CACCTAGGGGAGACACACCGCCTGTTTCCCAACACCATGTCTTGGCTCAGAGGCCCTG 1247

Db 1081 CACCTAGGGGAGACACACCGCCTGTTTCCCAACACCATGTCTTGGCTCAGAGGCCCTG 1140

QY 1248 TGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGTAGGCTAGGCTCCTGGGATCGAGGGATGCA 1307

Db 1141 TGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGTAGGCTCCTGGGATCGAGGGATGCA 1200

QY 1308 GTACAGCCACAGCATCATCAGAACCTTCTGTACCATGTGGTGGCTGGACCGACTGGAA 1367

Db 1201 GTACAGCCACAGCATCATCAGAACCTTCTGTACCATGTGGTGGCTGGACCGACTGGAA 1260

QY 1368 CCTTGGCTGAAACCCCGAAGGAGGAGCCCAATTTGGTGGCTAACTTTGTGACAGTCCCAT 1427

Db 1261 CCTTGGCTGAAACCCCGAAGGAGGAGCCCAATTTGGTGGCTAACTTTGTGACAGTCCCAT 1320

QY 1428 CATTGTAGACATCACCAAGGAGACGTTTACAAACAGCCCATGTTCTTACCACTTGGCCA 1487

Db 1321 CATTGTAGACATCACCAAGGAGACGTTTACAAACAGCCCATGTTCTTACCACTTGGCCA 1380

QY 1488 CTTTACGAAAGTTTCACTTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAGAAAG 1547

Db 1381 CTTTACGAAAGTTTCACTTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTCAGAAAG 1440

QY 1548 CGACCTGGACGCGAGTGGCAGTGTGATGATCCCGATGGCTTCTGCTGTTGTTGCTGCTAAA 1607

Db 1441 CGACCTGGACGCGAGTGGCAGTGTGATGATCCCGATGGCTTCTGCTGTTGTTGCTGCTAAA 1500

QY 1608 CCGCTCTCTAAGGATGTGCTTACCATCAAGGATCCTGCTGTTGGGCTTCTTGGAGAC 1667

Db 1501 CCGCTCTCTAAGGATGTGCTTACCATCAAGGATCCTGCTGTTGGGCTTCTTGGAGAC 1560

QY 1668 AATCTCACCTGGCTTCTTCACTTACACACTTACCTGTGGGCTGGCCAGTGTGAGCAGATA 1727

Db 1561 AATCTCACCTGGCTTCTTCACTTACACACTTACCTGTGGGCTGGCCAGTGTGAGCAGATA 1620

QY 1728 CTCAGAGGAGGACTGGGCTCAGCCTTGGGCTTAAAGGAGACA 1768

Db 1621 CTCAGAGGAGGACTGGGCTCAGCCTTGGGCTTAAAGGAGACA 1661

AX147658
LOCUS AX147658 1611 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 7 from Patent WO0136603.
ACCESSION AX147658
VERSION AX147658.1 GI:14346714
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1611)
Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.
Recombinant adeno-associated virus virions for the treatment of
lyosomal disorders
Patent: WO 0136603-A 7 25-MAY-2001;
Avigen, Inc. (US); Children's Hospital Medical Center of Northern
California (US)
FEATURES
source Location/Qualifiers
1..1611
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1611
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41173.1"
/db_xref="GI:14346715"
/translation="MEFSSPSRECEPKPLNDPEGMAGSLTGLLLQAVSWASGARPCI
PKSFGYSVVCNATYCDSPPTFPALGTFSESTRSRRMELSMGPQIANHTGT
GLLTLQPEQKFKQVKGFGAMTDAALNILALSPQNLKLSYFSEEGIGNIIRV
PMASCDFSIRTYADTPDDQLHNSLPEEDTKLPIHLRALQAPVSLASPW
TSPTWLKTNGAVNGKSLKGPDIYHOTWARYFKFLDAYAEHLKQFWAVTAENEPS
AGLSGYPFQCLGTFPEHORDFIARDLPTLANSTHNRLMLDDQRLPHWAKVV
LTDPEAAKYVHGIADVHWDLPAPAKATLGETHRLFPNTMLFASEACVGSFWEQSVR
LGSWDRGMOYSHSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDITKDTF
YKQPMFYHLGHFSKFIPEGSORVGLVASQKNDLDAVLMHPDGSVVVVVNLRRSKDVP
LTKDPAVGFLFTISPGYSIHTYLWHRQ"
BASE COUNT 359 a 487 c 404 g 361 t
ORIGIN

Query Match 69.6%; Score 1587; DB 6; Length 1611;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1596; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 106 ATGGAGTTTCAAGTCTCCAGAGAGGAAATGTCCTCCCAAGCCCTTTGAGTAGGGTAAGCATC 165
Db 1 ATGGAGTTTCAAGTCTCCAGAGAGGAAATGTCCTCCCAAGCCCTTTGAGTAGGGTGGGG 60
QY 166 ATGGCTGGCAGCCCTACAGGATTGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 225
Db 61 ATGGCTGGCAGCCCTACAGGTTTGTCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 120
QY 226 CGCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 285
Db 121 CGCCCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 180
QY 286 TACTGTGACTCCTTTGACCCCCCGACCTTTCTGCTCTCAAGATCCTTGGTACCTCAGCCGCTATGAG 345
Db 181 TACTGTGACTCCTTTGACCCCCCGACCTTTCTGCTCTCAAGATCCTTGGTACCTCAGCCGCTATGAG 240
QY 346 AGTACACGAGTGGGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACG 405
Db 241 AGTACACGAGTGGGACGGATGGAGCTGAGTATGGGCCCCATCCAGGCTAATCACACG 300
QY 406 GGCACAGCCCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCCAGAAAGTGAAGGGATT 465
Db 301 GGCACAGCCCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCCAGAAAGTGAAGGGATT 360
QY 466 GGAGGGCCATGACAGATGCTGCTCTCAAGATCCTTGGCCCTGTACCCCTGCCCCAA 525
Db 361 GGAGGGCCATGACAGATGCTGCTCTCAAGATCCTTGGCCCTGTACCCCTGCCCCAA 420
QY 526 AATTGCTACTTAATCGTACTTCTCTGAAGAGGAATCGGATATACATCATCCGGGTA 585
Db 1501 AACCCTCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTCTGGAG 1560

Db 421 AATTGCTACTTAATCGTACTTCTCTGAAGAGGAATCGGATATAACATCATCCGGGTA 480
QY 586 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGAT 645
Db 481 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGTATGAT 540
QY 646 TCCAGTTGCACAACCTCAGCCTCCAGAGGAGAGATACCAAGCTCAAGATACCCCTGATT 705
Db 541 TCCAGTTGCACAACCTCAGCCTCCAGAGGAGAGATACCAAGCTCAAGATACCCCTGATT 600
QY 706 CACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTTCCAGCCCTGGACATCA 765
Db 601 CACCGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTTCCAGCCCTGGACATCA 660
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 825
Db 661 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 720
QY 826 GGAGACATCTACCACCAAGCTGGGCCAGATACCTTGTGAAGTTCTCTGGATGCCCTATGCT 885
Db 721 GGAGACATCTACCACCAAGCTGGGCCAGATACCTTGTGAAGTTCTCTGGATGCCCTATGCT 780
QY 886 GACCACAAGTTACAGTTCTGGCCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 945
Db 781 GACCACAAGTTACAGTTCTGGCCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 840
QY 946 AGTGGATACCCCTTCCAGTGCCTGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCC 1005
Db 841 AGTGGATACCCCTTCCAGTGCCTGGCTTCCACCCCTGAACATCAGCGAGACTTCATTGCC 900
QY 1006 CGTGACCTAGGTCTTACCTCGCCCAACAGTACTACCAACATGTCCGCCCTACTCATGCTG 1065
Db 901 CGTGACCTAGGTCTTACCTCGCCCAACAGTACTACCAACATGTCCGCCCTACTCATGCTG 960
QY 1066 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
Db 961 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1126 GCTAAATATGTTTATGGCATTTGCTGTACATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1185
Db 1021 GCTAAATATGTTTATGGCATTTGCTGTACATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
QY 1186 GCCACCTAGGGGAGACACACCGCCTGTTCCCCCAACACCATGCTCTTTGCTGCTGAGGCCC 1245
Db 1081 GCCACCTAGGGGAGACACACCGCCTGTTCCCCCAACACCATGCTCTTTGCTGCTGAGGCCC 1140
QY 1246 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1305
Db 1141 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1200
QY 1306 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGG 1365
Db 1201 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGG 1260
QY 1366 AACCTTGGCCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCTAACTTTGTGCGACAGTCCC 1425
Db 1261 AACCTTGGCCCTGAACCCCGAAGGAGGACCCCAATTTGGTGGCTAACTTTGTGCGACAGTCCC 1320
QY 1426 ATCATTTGTAGACATCACCAAGGACACAGTCTTACAAACAGCCCATGTTCTACCACTTGGC 1485
Db 1321 ATCATTTGTAGACATCACCAAGGACACAGTCTTACAAACAGCCCATGTTCTACCACTTGGC 1380
QY 1486 CACTTCAGCAAGTTTCATTTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTACAGAAG 1545
Db 1381 CACTTCAGCAAGTTTCATTTCTGAGGGCTCCAGAGAGTGGGGCTGGTTGCCAGTACAGAAG 1440
QY 1546 AACGACCTGGACGACGAGTGGCCTGATGATCATCCCGATGCTGCTGTGTGTGGTCTGCTA 1605
Db 1441 AACGACCTGGACGACGAGTGGCCTGATGATCATCCCGATGCTGCTGTGTGTGGTCTGCTA 1500
QY 1606 AACCGCTCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTCTGGAG 1665
Db 1501 AACCGCTCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTCTGGAG 1560

QY 1666 ACAATCTCACCTGGCTACTCCATTACACACCTACCTGTGGCGTGGCCAGTGA 1716
|||||
Db 1561 ACAATCTCACCTGGCTACTCCATTACACACCTACCTGTGGCATGCCAGTGA 1611
|||||
RESULT 8
AX147656
LOCUS AX147656 1551 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0136603.
ACCESSION AX147656
VERSION AX147656.1 GI:14346712
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.
TITLE Recombinant adeno-associated virus virions for the treatment of
lyosomal disorders
JOURNAL Patent: WO 0136603-A 5 25-MAY-2001;
Avigen, Inc. (US) ; Children's Hospital Medical Center of Northern
California (US)
FEATURES
source Location/Qualifiers
1..1551
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1551
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41172.1"
/db_xref="GI:14346713"
translation="MAGSLTGLLLQAVSWASGARPCIPKSFYSSVVCNATYGDS
FDPTFPALCTFSYESTRSGRRMELSMGPIQANTGTGLLLTLQPEQKFKVKGFGG
AMTDAALNIALPQAQNLKLSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDD
FOLHNFSLPEEDTKLIPLIHRLAQAPVSLASPTWTLKTNAGAVNGKSLKG
QFGDIYHQTWARVFKFLDVAEHLQFVAVTAENEPSAGLLSGYFPQCLGFTPEHQR
DFARDLGPPTLANSTHNVNLLMDLQRLLLPHWAKVLTDPAAKYVHGIADVHVLVD
FLAPAKATLGETHRLFPNTMLFASACVSGKFWESVRLGSDWDRGMQYSHSIITNLLY
HVVGWTDNLANLPEGGPNWVRNFVDSPIIVDTKDTFYKQPMFYHLGHFSKFIPEG
ORVGLVASQKNDLDAVALMHPDGSVVVWLNRRSKDVPLTIKDPVAVGELETISPGYSI
HTYLWHRQ"
BASE COUNT 344 a 474 c 387 g 346 t
ORIGIN
Query Match 67.9%; Score 1547.8; DB 6; Length 1551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 166 ATGGCTGGCAGCCCTACAGGATGCTTCTACTTCAGGCAGTGTCTGGGCATCAGGTGCC 225
|||||
Db 1 ATGGCTGGCAGCCCTACAGGTTTCTTCTACTTCAGGCAGTGTCTGGGCATCAGGTGCC 60
QY 226 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACA 285
|||||
Db 61 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGCAATGCCACA 120
QY 286 TACTGTGACTCCTTTGACCCCGACCGGCTGAGTATGGTGGTGGTGGTGGTGGTGGTGGT 345
|||||
Db 121 TACTGTGACTCCTTTGACCCCGACCGCTTCTCTGCTTGGTACCTTCAGCCGCTATGAG 180
QY 346 AGTACACGCAGTGGCGACGGATGAGCTGAGTATGGGCGCCATCCAGGCTAATCACACG 405
|||||
Db 181 AGTACACGCAGTGGCGACGGATGAGCTGAGTATGGGCGCCATCCAGGCTAATCACACG 240
QY 406 GGACAGGCCCTGCTACTGACCCCTGACGCCAGCAAGAAAGTTCCAGAAAGTGAAGGATTT 465
Db 241 GGACAGGCCCTGCTACTGACCCCTGACGCCAGCAAGAAAGTTCCAGAAAGTGAAGGATTT 300
QY 466 GGAGGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCCTGTACCCCTGCCCCAA 525
|||||
Db 301 GGAGGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCCTGTACCCCTGCCCCAA 360
|||||

QY 526 AATTGCTACTTAATCGTACTTCTCTGAAGAAGAAATCGGATATAAATCATCATCCGGGTA 585
|||||
Db 361 AATTGCTACTTAATCGTACTTCTCTGAAGAAGAAATCGGATATAAATCATCATCCGGGTA 420
|||||
QY 586 CCCATGGCCAGCTGTGACTTCTCCATCGGCACCTACACCTATGCAGACACCCCTGATGAT 645
|||||
Db 421 CCCATGGCCAGCTGTGACTTCTCCATCGGCACCTACACCTATGCAGACACCCCTGATGAT 480
|||||
QY 646 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
|||||
Db 481 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540
|||||
QY 706 CACCGAGCCCTGCAGTTGGCCCGCAGCTCCCGTTTCTACTCTTCCAGCCCTGGACATCA 765
|||||
Db 541 CACCGAGCCCTGCAGTTGGCCCGCAGCTCCCGTTTCTACTCTTCCAGCCCTGGACATCA 600
|||||
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCTACTCAAGGGACAGCCC 825
|||||
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCTACTCAAGGGACAGCCC 660
|||||
QY 826 GGAGACATCTACCACACAGACCTGGGCCAGATACCTTTGTGAAGTTCTCTGGATGCCATGCT 885
|||||
Db 661 GGAGACATCTACCACACAGACCTGGGCCAGATACCTTTGTGAAGTTCTCTGGATGCCATGCT 720
|||||
QY 886 GAGCACAAAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGTGGCTGTG 945
|||||
Db 721 GAGCACAAAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGTGGCTGTG 780
|||||
QY 946 AGTGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCC 1005
|||||
Db 781 AGTGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCC 840
|||||
QY 1006 CGTGACTAGGTCTTACCTCGCCACAGTACTCACCACAAATGTCCGCTTACTCATGCTG 1065
|||||
Db 841 CGTGACTAGGTCTTACCTCGCCACAGTACTCACCACAAATGTCCGCTTACTCATGCTG 900
|||||
QY 1066 GATGACAAAGCTTGTGCTGCTGCCCTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125
|||||
Db 901 GATGACAAAGCTTGTGCTGCTGCCCTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
|||||
QY 1126 GCTAAATATGTTTCATGGCATGCTGTGATACATTTGTTACCTGGACTTTCTGGCTCCAGCCAAA 1185
|||||
Db 961 GCTAAATATGTTTCATGGCATGCTGTGATACATTTGTTACCTGGACTTTCTGGCTCCAGCCAAA 1020
|||||
QY 1186 GCCACCTAGGGGAGACACACCGCTGTTCCCAACACCATGCTCTTTCCTTCAGAGGCC 1245
|||||
Db 1021 GCCACCTAGGGGAGACACACCGCTGTTCCCAACACCATGCTCTTTCCTTCAGAGGCC 1080
|||||
QY 1246 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1305
|||||
Db 1081 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATG 1140
|||||
QY 1306 CAGTACAGCCACAGCATCATCAGAACTTCTGTACCATGTGGTGGCTGGACCGACTGG 1365
|||||
Db 1141 CAGTACAGCCACAGCATCATCAGAACTTCTGTACCATGTGGTGGCTGGACCGACTGG 1200
|||||
QY 1366 AACCTTGGCTGAAACCCGAGAGGAGACCCCAATTGGGTGCGTAACCTTGTTCGACAGTCCC 1425
|||||
Db 1201 AACCTTGGCTGAAACCCGAGAGGAGACCCCAATTGGGTGCGTAACCTTGTTCGACAGTCCC 1260
|||||
QY 1426 ATCATTTAGACATCAACCAAGGACACGTTTACAAACAGCCCATGTTTCTACCACCTTGGC 1485
|||||
Db 1261 ATCATTTAGACATCAACCAAGGACACGTTTACAAACAGCCCATGTTTCTACCACCTTGGC 1320
|||||
QY 1486 CACTTCAGCAAGTTTCATCTCTGAGGGTCCCAGAGAGTGGGCTGGTTCAGTCAAGAG 1545
|||||
Db 1321 CACTTCAGCAAGTTTCATCTCTGAGGGTCCCAGAGAGTGGGCTGGTTCAGTCAAGAG 1380
|||||
QY 1546 AACGACCTGGACCGCAGTGGCAGTATGATCATCCCGATGGCTCTGCTGTGTGGTCTGCTA 1605
|||||
Db 1381 AACGACCTGGACCGCAGTGGCAGTATGATCATCCCGATGGCTCTGCTGTGTGGTCTGCTA 1440
|||||

QY 1606 AACCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGCTTCTCTGGAG 1665
|||||
Db 1441 AACCGCTCCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGCTTCTCTGGAG 1500
|||||
QY 1666 ACAATCTCACCTGGCTACTCCATTACACCTACCTGTGGCGTCCAGTGA 1716
|||||
Db 1501 ACAATCTCACCTGGCTACTCCATTACACCTACCTGTGGCATCCAGTGA 1551
|||||

RESULT 9
HUMGCB A HUMGCB A 1823 bp mRNA linear PRI 03-FEB-1999
LOCUS Human mRNA for alternative splicing product of glucocerebrosidase
DEFINITION pseudogene, complete cds.
ACCESSION D13287
VERSION D13287.1 GI:496370
KEYWORDS alternative splicing product; glucocerebrosidase; pseudogene.
SOURCE Homo sapiens (library: lambda hGC12) cell-line HL-60 cDNA to mRNA,
clone pGC12A.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1823)
AUTHORS Imai,K., Nakamura,M., Yamada,M., Asano,A., Yokoyama,S., Tsuji,S.
and Ginns,E.I.
TITLE A novel transcript from a pseudogene for human glucocerebrosidase
in non-Gaucher disease cells
JOURNAL Gene 136 (1-2), 365-368 (1993)
MEDLINE 94124033
REFERENCE 2 (bases 1 to 1823)
AUTHORS Nakamura,M.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1992) Masahiko Nakamura, Osaka University,
Institute for Protein Research; 3-2 Yamadaoka, Suita, Osaka 565,
Japan (E-mail:masahiko@protein.osaka-u.ac.jp, Tel:06-879-8628,
Fax:06-879-8629)
COMMENT On Jun 5, 1994 this sequence version replaced gi:441043.
FEATURES
Location/Qualifiers
source 1..1823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HL-60"
/clone_lib="lambda hGC12"
209..1000
/note="alternative splicing product of glucocerebrosidase
pseudogene"
/codon_start=1
/product="glucocerebrosidase"
/protein_id="BAA02546.1"
/db_xref="GI:2160388"
/translation="MASCDFSIRTYTYADTPDFQLHNFSLPEEDTKLKIPLIHLALQ
LAQRPVSLASPWTSPRLKTRGAGNGKPLKQOPRDIYHOTWARYIVKFLDAYAEHK
LQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLPTLANGTHHVRLLMLD
DORLLPHWAKVLTDPAAKYVHGIADVWLYDLAPAKATLRETHLPNTMLFASE
ACVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDWNPSL"
polyA_site 1823
BASE COUNT 435 a 525 c 489 g 374 t
ORIGIN
Query Match 67.8%; Score 1544.2; DB 9; Length 1823;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 23; Indels 55; Gaps 1;

QY 558 AGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCAC 617
|||||
Db 178 AGGAATCGGATATAACATCATCTGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCAC 237
|||||
QY 618 CTACACCTATGACAGACACCCCTGATGATTCCAGTTGCACAACTTCAGCCTCCAGAGGA 677
|||||
Db 238 CTACACCTATGACAGACACCCCTGATGATTCCAGTTGCACAACTTCAGCCTCCAGAGGA 297
|||||
QY 678 AGATACCAAGCTCAAGATACCCCTGATTACCGAGCCCTGCAGTTGGCCCGCAGCGTCCCGT 737
|||||

Db 298 AGATACCAAGCTCAAGATACCCCTGATTTCACCGAGCCCTGCAGTTGGCCCGCAGCGTCCCGT 357
QY 738 TTCACCTCCTTGCCAGCCCTGGACATCACCCCACTTGGCTCAAGACCAATGGAGCGGTGAA 797
|||||
Db 358 TTCACCTCCTTGCCAGCCCTGGACATCACCCCACTTGGCTCAAGACCAAGGAGCGGGGAA 417
|||||
QY 798 TGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACAGACTTGGGGCCAGATA 857
|||||
Db 418 TGGGAAGGGGCCACTCAAGGGACAGCCCGGAGACATCTACCACAGACTTGGGGCCAGATA 477
|||||
QY 858 CTTTGTGAAGTTCTTGGATGCCCTATGCTGAGCAACAAGTTACAGTTCTGGCAGTGACAGC 917
|||||
Db 478 CATTTGTGAAGTTCTTGGATGCCCTATGCTGAGCAACAAGTTACAGTTCTGGCAGTGACAGC 537
|||||
QY 918 TGAAAATGAGCCCTTCTGCTGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 977
|||||
Db 538 TGAAAATGAGCCCTTCTGCTGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 597
|||||
QY 978 CCCTGAACATCAGCGAGACTTCATTGCCCGTACCCTAGGTCCCTACCCCTGCCAACAGTAC 1037
|||||
Db 598 CCCTGAACATCAGCGAGACTTCATTGCCCGTACCCTAGGTCCCTACCCCTGCCAACAGTAC 657
|||||
QY 1038 TCACCACAAATGTCGCCCTACTCATGCTGGATGACCAACGCTTGTGCTGCCCACTGGGC 1097
|||||
Db 658 TCACCACAAATGTCGCCCTACTCATGCTGGATGACCAACGCTTGTGCTGCCCACTGGGC 717
|||||
QY 1098 AAAGGTGGTACTGACAGACCCAGAACGAGCTAAATATGTTTCATGGCATGCTGTACATTG 1157
|||||
Db 718 AAAGGTGGTACTGACAGACCCAGAACGAGCTAAATATGTTTCATGGTATTGTGTACATTG 777
|||||
QY 1158 GTACCTGGACTTTCTGCTGCCAGCCAAAGCCACCTAGGGGAGACACACCGCTGTTCCTC 1217
|||||
Db 778 GTACCTGGACTTTCTGCTGCCAGCCAAAGCCACCTAGGGGAGACACACCGCTGTTCCTC 837
|||||
QY 1218 CAACACCATGCTCTTTGCTCAGAGGCCCTGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGT 1277
|||||
Db 838 CAACACCATGCTCTTTGCTCAGAGGCCCTGTGTGGGTTCCAAGTTCTGGGAGCAGAGTGT 897
|||||
QY 1278 GCGGCTAGGCTCCTGGATCGAGGGATGCAGTACAGCCACAGCATCATCAGAAACCTCCT 1337
|||||
Db 898 GCGGCTAGGCTCCTGGATCGAGGGATGCAGTACAGCCACAGCATCATCAGAAACCTCCT 957
|||||
QY 1338 GTACCATGTGGTGGCTGGACCGACTTGAACCTTGCCTTGAACCCCGAAGGAGGACCCAA 1397
|||||
Db 958 GTACCATGTGGTGGCTGGACCGACTTGAACCTTGCCTTGAACCCCGAAGGAGGACCCAA 987
|||||
QY 1398 TTGGGTGCGTAACCTTTCGACAGTCCCATCATATGTAGACATCACCAAGGACACGTTT 1457
|||||
Db 988 -----CCCATCATATGTAGACATCACCAAGGACACACGTTT 1022
|||||
QY 1458 CAAACAGCCCATGTTCTACCACTTGGCCACTTCCAGCAAGTTTCATTCCTGAGGGCTCCCA 1517
|||||
Db 1023 CAAACAGCCCATGTTCTACCACTTGGCCACTTCCAGCAAGTTTCATTCCTGAGGGCTCCCA 1082
|||||
QY 1518 GAGAGTGGGGCTGGTTGCCAGTCAGAAAGAACGACCTGGACGCGAGTGGCACTGATGATCC 1577
|||||
Db 1083 GAGAGTGGGGCTGGTTGCCAGTCAGAAAGAACGACCGGACGCGAGTGGCACTGATGATCC 1142
|||||
QY 1578 CGATGGCTCTGCTGTTGGTCTGTCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1637
|||||
Db 1143 CGATGGCTCTGCTGTTGGTCTGTCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1202
|||||
QY 1638 CAAGGATCCTGCTGTGGCTTCTGGAGACAAATCTACCTGGCTACTCCATTCACACCTA 1697
|||||
Db 1203 CAAGGATCCTGCTGTGGCTTCTGGAGACAAATCTACCTGGCTACTCCATTCACACCTA 1262
|||||
QY 1698 CCTGTGGCGTCGCCAGTGGAGCAGATACTCAAGGAGGCGACTGGGCTCAGCCTGGGCA 1757
|||||
Db 1263 CCTGTGGCGTCGCCAGTGGAGCAGATACTCAAGGAGGCGACTGGGCTCAGCCTGGGCA 1322
|||||
QY 1758 TAAAGGGACAGAGTCAAGTCAAGCCTGTCTGTGACTAAAGAGGGGCGACAGCGGCCAG 1817
|||||
Db 1323 TAAAGGGACAGAGTCAAGTCAAGCCTGTCTGTGACTAAAGAGGGGCGACAGCGGCCAG 1382
|||||

QY 1818 TGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGGTTTGGGTGACTCACTTTCCCTCT 1877
Db 1383 CTTGACGTTACAGCGACGTAAGCCAGGGGCAATGGTTTGGGTGACTCACTTTCCCTCT 1442
QY 1878 AGGTGGTCCAGGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCAAGTGTCCCTCCA 1937
Db 1443 AGGCGGTGCCAGGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCAAGTGTCCCTCCA 1502
QY 1938 GCGCCCATGCTTATGTGAACATGGCTGTGTGCTGCTTGTCTTGGAAACTGGGCTGGGT 1997
Db 1503 GCGCCCATGCTTATGTGAACATGGCTGTGTGCTGCTTGTCTTGGAAACTGGGCTGGGT 1562
QY 1998 CCAGGCCCTAGGTGAGCTCACTGTCCGTACAAACACAAAGATCAGGGCTGAGGTAAGGAA 2057
Db 1563 CCAGGCCCTAGGTGAGCTCACTGTCCGTACAAACACAAAGATCAGGGCTGAGGTAAGGAA 1622
QY 2058 AAGAAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTTGTCTTTCTCTGGAGATGC 2117
Db 1623 AAGAAGAGACTAGGAAAGCTGGGCCCAAACTGGAGACTGTTTGTCTTTCTCTGGAGATGC 1682
QY 2118 AGAACTGGCCCTGGAGCAGCAGTGTCAAGCATCAGGGCGGGAAGCTTAAAGCAGCAGCG 2177
Db 1683 AGAACTGGCCCTGGAGCAGCAGTGTCAAGCATCAGGGCGGGAAGCTTAAAGCAGCAGCG 1742
QY 2178 GGTGTGCCCCAGGACCCAGATGATTCCTATGGCAGCCAGCCAGGAAAATGGCAGCTCTTA 2237
Db 1743 GGTGTGCCCCAGGACCCAGATGATTCCTATGGCAGCCAGCCAGGAAAATGGCAGCTCTTA 1802
QY 2238 AAGGAGAAAATGTTTGAGCCC 2258
Db 1803 AAGGAGAAAATGTTTGAGCCC 1823

RESULT 10
AX299779
LOCUS AX299779 1593 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 3 from Patent WO0177307.
ACCESSION AX299779
VERSION AX299779.1 GI:17129300
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Berent,S.L.
TITLE Expression system for efficiently producing clinically effective
lysosomal enzymes (glucocerebrosidase)
JOURNAL Patent: WO 0177307-A 3 18-OCT-2001;
CYTOKLONAL PHARMACEUTICS, INC. (US)
FEATURES
source location/Qualifiers
1..1593
/organism="Homo sapiens"
/db_xref="taxon:9606"
13..1563
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD12721.1"
/db_xref="GI:17129301"
/translation="MAGSLTGLLLQAVSWASGARPCIPKSFYSSVVCNATYCD
/translaton="MAGSLTGLLLQAVSWASGARPCIPKSFYSSVVCNATYCD
FDPPTPALGTFESRYESTRSGRMELSMGPIQANHTGTGLLLTLOPEQKQKVGFG
AMTDAALNIALSPAPQNLKLSYFSEEGIGNIIRVPMASCDPSIRTYADTPDD
FOLNFSLPEDTKLPIIHRALQAOQPVSLASPTSPWLKTNCAVNGKSLKG
QPGDIYHOTWARYEVKFLDAYAEHLQFVAWTAENEPSAGLLSGYFQCLGFTPEHQ
DFIARDLPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIHWHYLD
FLAPAKATLGETHRLFPNTMLFSEACVGSKFQSVRLGSDRGMOYSHSIITNLLY
HVYGLVDNWLALNPEGGPNWVNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPEG
QRYGLVASKNDLDAVALMHPDGSAAVVVNLNRSSKDVPPLTIKDPVAGFLETISPYSI
HTYLWHRQ"
BASE COUNT 359 a 484 c 396 g 354 t
ORIGIN

Query Match 67.7%; Score 1544; DB 6; Length 1593;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1550; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 157 GTAACCATCATGGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTCTGTGGCA 216
Db 4 GAATTCAAAATGGCTGGCAGCCTCACAGGTTTCTTCTACTTCAGGCAGTGTCTGTGGCA 63
QY 217 TCAGGTGCCCCCGCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGC 276
Db 64 TCAGGTGCCCCCGCTGCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTCTGC 123
QY 277 AATGCCACATATCTGTGACTCCTTTGACCCCCCGACCTTTCTGCTGCTTGGTACCTTCAGC 336
Db 124 AATGCCACATATCTGTGACTCCTTTGACCCCCCGACCTTTCTGCTGCTTGGTACCTTCAGC 183
QY 337 CGCTATGAGAGTACACGCGAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCT 396
Db 184 CGCTATGAGAGTACACGCGAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCT 243
QY 397 AATCACACGGGCACAGGCTGCTACTGACCTGCGAGCAGAACAGAGTTCAGAAAGTG 456
Db 244 AATCACACGGGCACAGGCTGCTACTGACCTGCGAGCAGAACAGAGTTCAGAAAGTG 303
QY 457 AAGGATTTGGAGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGGCCCTGTCAACC 516
Db 304 AAGGATTTGGAGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGGCCCTGTCAACC 363
QY 517 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 576
Db 364 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 423
QY 577 ATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACC 636
Db 424 ATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACC 483
QY 637 CCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 696
Db 484 CCTGATGATTTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 543
QY 697 CCCCTGATTCACCGAGCCCTGCAGTTGGCCCGCCAGCTCCCGTTCCTCTTGGCAGCCCT 756
Db 544 CCCCTGATTCACCGAGCCCTGCAGTTGGCCCGCCAGCTCCCGTTCCTCTTGGCAGCCCT 603
QY 757 TGGACATCACCCACCTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAG 816
Db 604 TGGACATCACCCACCTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAG 663
QY 817 GGACAGCCCGGAGACATCTACACAGACCTGGCCAGACATCTTTGTGAAGTTTCTGTGAT 876
Db 664 GGACAGCCCGGAGACATCTACACAGACCTGGCCAGACATCTTTGTGAAGTTTCTGTGAT 723
QY 877 GCCTATGCTGAGCACAAGTTACAGTTCTTGGCCAGTGCAGCTGAAATGAGCCCTTCTGCT 936
Db 724 GCCTATGCTGAGCACAAGTTACAGTTCTTGGCCAGTGCAGCTGAAATGAGCCCTTCTGCT 783
QY 937 GGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAAACATCAGCGAGAC 996
Db 784 GGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAAACATCAGCGAGAC 843
QY 997 TTCATGCTGCTGAGTACCTAGGTCTTACCTGCGCAACAGTACTACACAAATGTCGGCTA 1056
Db 844 TTCATGCTGCTGAGTACCTAGGTCTTACCTGCGCAACAGTACTACACAAATGTCGGCTA 903
QY 1057 CTCATGCTGATGACCAACGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1116
Db 904 CTCATGCTGATGACCAACGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 963
QY 1117 CCAGAGCAGCTAAATATGTTTCATGGCATGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCT 1176
Db 964 CCAGAGCAGCTAAATATGTTTCATGGCATGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCT 1023
QY 1177 CCAGCCAAAGCCACCTAGGGGAGACACACCCGCTGTCTCCCAACACCATGCTCTTTGCC 1236


```
|||||
Db 1024 CCAGCCAAAGCCACCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTTGCC 1083
QY 1237 TCAGAGGCTGTGGGCTCCAAGTTCTGGAGCAGAGTGTGCGGTAGGCTCCTGGGAT 1296
Db 1084 TCAGAGGCTGTGGGCTCCAAGTTCTGGAGCAGAGTGTGCGGTAGGCTCCTGGGAT 1143
QY 1297 CGAGGGATGCAGTACAGCCACAGCATCATCAGCAACCTCCTGTACATGTGGTGGCTGG 1356
Db 1144 CGAGGGATGCAGTACAGCCACAGCATCATCAGCAACCTCCTGTACATGTGGTGGCTGG 1203
QY 1357 ACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCCAATTGGGTGCGTAACCTTTGTC 1416
Db 1204 ACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCCAATTGGGTGCGTAACCTTTGTC 1263
QY 1417 GACAGTCCCACATTTAGACATCACCAGGACACAGTCTTACAAACAGCCCATGTTCTAC 1476
Db 1264 GACAGTCCCACATTTAGACATCACCAGGACACAGTCTTACAAACAGCCCATGTTCTAC 1323
QY 1477 CACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCCAGAGAGTGGGCTGGTGGC 1536
Db 1324 CACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCCAGAGAGTGGGCTGGTGGC 1383
QY 1537 AGTCAGAAGAACGACCTGGACGAGTGGGCTGATGATCCCGATCCGATGGCTCTGCTGTGTG 1596
Db 1384 AGTCAGAAGAACGACCTGGACGAGTGGGCTGATGATCCCGATCCGATGGCTCTGCTGTGTG 1443
QY 1597 GTCGTGCTAAACGCTCCTCTAAGGATGTCCTTACCATCAAGGATCCTGCTGTGGC 1656
Db 1444 GTCGTGCTAAACGCTCCTCTAAGGATGTCCTTACCATCAAGGATCCTGCTGTGGC 1503
QY 1657 TTCTGGAGACAACTTCACCTGGCTACTCCATTACACCTACCTGTGGCTGCGCCAGTGA 1716
Db 1504 TTCTGGAGACAACTTCACCTGGCTACTCCATTACACCTACCTGTGGCATAGACAATGA 1563

RESULT 11
AX191972
LOCUS AX191972 1551 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 2 from Patent WO0149830.
ACCESSION AX191972
VERSION AX191972.1 GI:15210120
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Okkels,J.S., Jensen,A.D., Halkier,T., Jensen,R.B. and Schambye,H.T.
TITLE Improved lysosomal enzymes and lysosomal enzyme activators
JOURNAL Patent: WO 0149830-A 2 12-JUL-2001;
Maxygen Aps (DK)
FEATURES
source
Location/Qualifiers
1..1551
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 345 a 470 c 389 g 347 t
ORIGIN
Query Match 67.7%; Score 1543; DB 6; Length 1551;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTGTTGGCATCAGGTGCC 225
Db 1 ATGGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTGTTGGCATCAGGTGCC 60
QY 226 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 285
Db 61 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 120
QY 286 TACTGTGATCCTTTGACCCCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAG 345
|||||
```

```
Db 121 TACTGTGACTCCTTTGACCCCCCGACCTTTCTCTGCCCTTGGTACCTTTCAGCCGCTATGAG 180
QY 346 AGTACACGCAGTGGCGGACGGATGGAGCTGAGTATGGGGCCCCATCCAGGCTAATCACACG 405
Db 181 AGTACACGCAGTGGCGGACGGATGGAGCTGAGTATGGGGCCCCATCCAGGCTAATCACACG 240
QY 406 GGCACAGGCCTGCTACTGACCCCTGCAGCCAGAGAACAGAGTTCCAGAAAGTGAAGGGATT 465
Db 241 GGCACAGGCCTGCTACTGACCCCTGCAGCCAGAGAACAGAGTTCCAGAAAGTGAAGGGATT 300
QY 466 GGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTGCCTGTACCCCTGCCCAA 525
Db 301 GGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTGCCTGTACCCCTGCCCAA 360
QY 526 AATTGTCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGTA 585
Db 361 AATTGTCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGTA 420
QY 586 CCCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGAT 645
Db 421 CCCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGCACAACCTTCAGCCTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
Db 481 TTCCAGTTGCACAACCTTCAGCCTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACCGAGCCTTCAGTTGGCCCCAGCGTCCCGTTTCACTCCTTGCAGCCCTGGACATCA 765
Db 541 CACCGAGCCTTCAGTTGGCCCCAGCGTCCCGTTTCACTCCTTGCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 825
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGGACAGCCC 660
QY 826 GGAGACATCTACCACGACCTTGGGCCAGATACCTTGTGAAGTTCTCTGGATGCCCTATGCT 885
Db 661 GGAGACATCTACCACGACCTTGGGCCAGATACCTTGTGAAGTTCTCTGGATGCCCTATGCT 720
QY 886 GAGCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 945
Db 721 GAGCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTTG 780
QY 946 AGTGGATACCCCTTCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTCATTGCC 1005
Db 781 AGTGGATACCCCTTCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGACTTAATTGCC 840
QY 1006 CGTGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACAATGTCGCTACTCATGCTG 1065
Db 841 CGTGACCTAGGTCTTACCTCGCCAAACAGTACTACCAACAATGTCGCTACTCATGCTG 900
QY 1066 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125
Db 901 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 1126 GCTAAATATGTTTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1185
Db 961 GCTAAATATGTTTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1020
QY 1186 GCCACCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTTCCTCAGAGGCC 1245
Db 1021 GCCACCTAGGGGAGACACACCGCCTGTTCCCAACACCATGCTCTTTCCTCAGAGGCC 1080
QY 1246 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGATCGAGGGATG 1305
Db 1081 TGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGATCGAGGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGCTGGACCGACTGG 1200
QY 1366 AACCTTGGCCTGAACCCCGAAGGAGGAGGCCCAATTTGGGTGCGTAACCTTGTGACAGTCCC 1425
Db 1201 AACCTTGGCCTGAACCCCGAAGGAGGAGGCCCAATTTGGGTGCGTAACCTTGTGACAGTCCC 1260
```

QY 1426 ATCATTTAGACATCAACAAGGACACAGTCTTACAAAACAGCCCATGTTCTACCACTTGGC 1485
Db 1261 ATCATTTAGACATCAACAAGGACACAGTCTTACAAAACAGCCCATGTTCTACCACTTGGC 1320
QY 1486 CACTTCAGCAAGTTCAATCTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCAGTCAGAAG 1545
Db 1321 CATTTAGCAAGTTCAATCTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCAGTCAGAAG 1380
QY 1546 AACGACCTGGACGAGTGGCACTGATGATCCCGATCCCGATGGCTGCTGTTGGTGGTCTGCTA 1605
Db 1381 AACGACCTGGACGAGTGGCACTGATGATCCCGATCCCGATGGCTGCTGTTGGTGGTCTGCTA 1440
QY 1606 AACCGCTCCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTTGGCTTCCCTGGAG 1665
Db 1441 AACCGCTCCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTTGGCTTCCCTGGAG 1500
QY 1666 ACAATCTACCTGGCTGACTCCATTCACACCTACCTGTGGCGTGGCCAGTGA 1716
Db 1501 ACAATCTACCTGGCTGACTCCATTCACACCTACCTGTGGCGTGGCCAGTGA 1551

RESULT 12
AX349364
LOCUS AX349364
DEFINITION Sequence 2 from Patent WO0202597.
ACCESSION AX349364
VERSION AX349364.1 GI:18615327
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE Okkels, J.S., Jensen, A.D. and van den Hazel, B.C.
AUTHORS Peptide extended glycosylated polypeptides
TITLE Patent: WO 0202597-A 2 10-JAN-2002;
JOURNAL Maxygen Aps (DK); Maxygen Holdings Ltd (KY)
FEATURES Location/Qualifiers
source 1..1551
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 345 a 470 c 389 g 347 t
ORIGIN

Query Match 67.7%; Score 1543; DB 6; Length 1551;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGCGTGGCAGCCTCACAGGATGCTTCTACTTACAGGAGTGTGCTGGGCATCAGGTGCC 225
Db 1 ATGCGTGGCAGCCTCACAGGATGCTTCTACTTACAGGAGTGTGCTGGGCATCAGGTGCC 60
QY 226 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGCTGTGCAATGCCACA 285
Db 61 CGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGCTGTGCAATGCCACA 120
QY 286 TACTGTGACTCCTTTGACCCCGGACCTTTCTGCTGCTTGGTACCTTCAGCCGCTATGAG 345
Db 121 TACTGTGACTCCTTTGACCCCGGACCTTTCTGCTGCTTGGTACCTTCAGCCGCTATGAG 180
QY 346 AGTACACGCGAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAG 405
Db 181 AGTACACGCGAGTGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAG 240
QY 406 GGCACAGGCTGCTACTGACCTTCAGCCCGCAGCAAGAGTTCAGAAAGTGAAGGATTT 465
Db 241 GGCACAGGCTGCTACTGACCTTCAGCCCGCAGCAAGAGTTCAGAAAGTGAAGGATTT 300
QY 466 GGAGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGCCCTGTACCCCTGCCCCAA 525
Db 301 GGAGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGCCCTGTACCCCTGCCCCAA 360

QY 526 AATTTGCTACTTAAATCGTACTTCTCTGAGAAGGAATCGGATATAACATCATCCGGGTA 585
Db 361 AATTTGCTACTTAAATCGTACTTCTCTGAGAAGGAATCGGATATAACATCATCCGGGTA 420
QY 586 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGAT 645
Db 421 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
Db 481 TTCCAGTTGCACAACTTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACCGAGCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCCTTGCCAGCCCTGGACATCA 765
Db 541 CACCGAGCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCCTTGCCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCCC 825
Db 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCCC 660
QY 826 GGAGACATCTACCACCCAGACCTGGGCCAGATACCTTTGTGAAGTTCTCGGATGCCATGCT 885
Db 661 GGAGACATCTACCACCCAGACCTGGGCCAGATACCTTTGTGAAGTTCTCGGATGCCATGCT 720
QY 886 GAGCACAAAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGCTGTTG 945
Db 721 GAGCACAAAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGCTGTTG 780
QY 946 AGTGATACCCCTTCCAGTGCCTGGCTTCAACCTGAAATCAGCGAGACTTCAATTGCC 1005
Db 781 AGTGATACCCCTTCCAGTGCCTGGCTTCAACCTGAAATCAGCGAGACTTCAATTGCC 840
QY 1006 CGTGACCTAGGTCTTACCTCGCCAAACAGTACTCACCACAATGTCCGCTACTCATGCTG 1065
Db 841 CGTGACCTAGGTCTTACCTCGCCAAACAGTACTCACCACAATGTCCGCTACTCATGCTG 900
QY 1066 GATGACCAACGCTTGTGCTGCCCTGCGCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 1125
Db 901 GATGACCAACGCTTGTGCTGCCCTGCGCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 960
QY 1126 GCTAAATATGTTTCATGSCATTGCTGTACATTTGGTACCTGGACTTTCTGGCTCCAGCCAAA 1185
Db 961 GCTAAATATGTTTCATGSCATTGCTGTACATTTGGTACCTGGACTTTCTGGCTCCAGCCAAA 1020
QY 1186 GCCACCTAGGGGAGACACACCGCTTCTCCCAACACCATGCTCTTTGCTCAGAGGCC 1245
Db 1021 GCCACCTAGGGGAGACACACCGCTTCTCCCAACACCATGCTCTTTGCTCAGAGGCC 1080
QY 1246 TGTGTGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGTAGGCTCCTGGGATCGAGGGATG 1305
Db 1081 TGTGTGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGTAGGCTCCTGGGATCGAGGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCAGAACTCCTGTACCATGTGGTCCGCTGGACCCGACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGAACTCCTGTACCATGTGGTCCGCTGGACCCGACTGG 1200
QY 1366 AACCTTGGCTGAAACCCCGAAGGAGGACCCAAATTGGGTGCGTAACCTTGTTCGACAGTCCC 1425
Db 1201 AACCTTGGCTGAAACCCCGAAGGAGGACCCAAATTGGGTGCGTAACCTTGTTCGACAGTCCC 1260
QY 1426 ATCATTTAGACATCAACAAGGACACAGTCTTACAAAACAGCCCATGTTCTACCACTTGGC 1485
Db 1261 ATCATTTAGACATCAACAAGGACACAGTCTTACAAAACAGCCCATGTTCTACCACTTGGC 1320
QY 1486 CACTTCAGCAAGTTCAATCTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCAGTCAGAAG 1545
Db 1321 CACTTCAGCAAGTTCAATCTCCTGAGGGCTCCAGAGAGTGGGGCTGGTGGCAGTCAGAAG 1380
QY 1546 AACGACCTGGACGAGTGGCAGTATGATCCCGATCCCGATGGCTGCTGCTGTTGGTGGTCTA 1605
Db 1381 AACGACCTGGACGAGTGGCAGTATGATCCCGATCCCGATGGCTGCTGCTGTTGGTGGTCTA 1440
QY 1606 AACCGCTCCTCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGGCGTTCCTGGAG 1665


```
|||||
Db 1441 AACCGCTCCTAAGGATGTCCTCTTACCATCAAGGATCCTGCTGGGCTTCTCTGGAG 1500
QY 1666 ACAATCTCACCTGGCTACTCCATTACACCTACCTGTGGCGTCGCCAGTGA 1716
Db 1501 ACAATCTCACCTGGCTACTCCATTACACCTACCTGTGGCGTCGCCAGTGA 1551

RESULT 13
AX299777 LOCUS AX299777 1592 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 1 from Patent WO0177307.
ACCESSION AX299777
VERSION AX299777.1 GI:17129298
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Berent,S.L.
TITLE Expression system for efficiently producing clinically effective
lysosomal enzymes (glucocerebrosidase)
JOURNAL Patent: WO 0177307-A 1 18-OCT-2001;
CYTOCLONAL PHARMACEUTICS, INC. (US)
FEATURES
source Location/Qualifiers
1..1592
/organism="Homo sapiens"
/db_xref="taxon:9606"
13..1563
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD12720.1"
/db_xref="GI:17129298"
/translation="MAGSLTGLLLQLQAVSWAGARPCIPKSPFGYSSVVCVNCATYCD
SDS FPPFPALGTFSTRSGRMELSMGPIQANHTGTLGLLLTLQPEQKFKYKFGGG
AMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDIFSIRTYTYADTPDD
FQLHNFSPLEEDTKLPIHRAQLAQRPVSLLSAPWTSPTWLKNGAVNGKSLKG
OPGDIYHQTWARYFVKFLDAYAEHLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHOR
DFIARDLGPTLANSTHNVRLMDQDQLLLPHWAKVYVLTDPAAKYVHVHGIYHLYLD
FLAPAKATLGETHRLFPNTMLFASACVGSKEQSVRLGSDRGMQYSHSIITNLLY
HVVGWTDNWLALNPEGGPNWRFNFDSPIIIVDITKDFYKQPMFYHLGHFSKFTPEGS
QRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAGFLETISPGYSI
HTYLWRRQ"
BASE COUNT 359 a 481 c 397 g 355 t
ORIGIN
Query Match 67.1%; Score 1529.6; DB 6; Length 1592;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1541; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 157 GTAAGCATCATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTGCGGGCA 216
Db 4 GAATCAAAATGGCTGGCTGCTTAACAGGATGCTTCTACTTCAGGCAGTGTGCGGGCA 63
QY 217 TCAGGTGCGCCCGCTGCATCCCTAAAGCTTCGGGTACAGCTCGGTGTGTGTCTGC 276
Db 64 TCCGGAGCTAGACCTTGATCCCTTAAAGCTTCGGGTACAGCTCGGTGTGTGTCTGC 123
QY 277 AATGCCACATACGTGACTCCTTTGACCCCGCCGACCTTCTGCTGCTGCTTACCTCAGC 336
Db 124 AATGCCACATACGTGACTCCTTTGACCCCGCCGACCTTCTGCTGCTGCTTACCTCAGC 183
QY 337 CGCTATGAGAGTACACGCGAGTGGCGGACGAGTGGAGTGGGCGCCATCCAGGCT 396
Db 184 CGCTATGAGAGTACACGCGAGTGGCGGACGAGTGGAGTGGGCGCCATCCAGGCT 243
QY 397 AATCACACGGGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAAAGTG 456
Db 244 AATCACACGGGCACAGGCTGCTACTGACCCCTGCAGCCAGAACAGAGTTCAGAAAGTG 303
QY 457 AAGGGATTGGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTGGCCTGTACCC 516
|||||
```

```
Db 304 AAGGGATTGGAGGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGGCCTGTACCC 363
QY 517 CTGCCCCAAAATTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 576
Db 364 CTGCCCCAAAATTTGCTACTTAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 423
QY 577 ATCCGGGTACCCATGGCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACC 636
Db 424 ATCCGGGTACCCATGGCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACC 483
QY 637 CCTGATGATTTCCAGTTGCACAACTTCCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 696
Db 484 CCTGATGATTTCCAGTTGCACAACTTCCAGCCTCCAGAGGAAGATACCAAGCTCAAGATA 543
QY 697 CCCCTGATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCCTTGCAGCCCC 756
Db 544 CCCCTGATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCCGTTTCACTCCTTGCAGCCCC 603
QY 757 TGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAG 816
Db 604 TGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAG 663
QY 817 GGACAGCCCGGAGACATCTACCACAGACCTGGGCCAGATACTTTGTGAAGTCTCTGGAT 876
Db 664 GGACAGCCCGGAGACATCTACCACAGACCTGGGCCAGATACTTTGTGAAGTCTCTGGAT 723
QY 877 GCCTATGCTGAGCACAAAGTTACAGTTCTGGGCAGTGCAGCTGAAATGAGCCTTCTGCT 936
Db 724 GCCTATGCTGAGCACAAAGTTACAGTTCTGGGCAGTGCAGCTGAAATGAGCCTTCTGCT 783
QY 937 GGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGAC 996
Db 784 GGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGGCTTACCCCTGAACATCAGCGAGAC 843
QY 997 TTCAATGCTGCTGACCTAGGTGCTACCTCGCCCAACAGTACTCACCACAATGTCCGCCTA 1056
Db 844 TTCAATGCTGCTGACCTAGGTGCTACCTCGCCCAACAGTACTCACCACAATGTCCGCCTA 903
QY 1057 CTCATGCTGGATGACCAACGCTGCTGCTGCCCCACCTGGSCAAAGGTGGTACTGACAGAC 1116
Db 904 CTCATGCTGGATGACCAACGCTGCTGCTGCCCCACCTGGSCAAAGGTGGTACTGACAGAC 963
QY 1117 CCAGAACGAGCTAAATATGTTTATGGCATGCTGTATGCTGCTGCTGCTGCTGCTGCT 1176
Db 964 CCAGAACGAGCTAAATATGTTTATGGCATGCTGTATGCTGCTGCTGCTGCTGCTGCT 1023
QY 1177 CCAGCCAAAGCCACCCTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTCTTGCC 1236
Db 1024 CCAGCCAAAGCCACCCTAGGGGAGACACACCCGCTGTTCCCAACACCATGCTCTTGCC 1083
QY 1237 TCAGAGGCTGTGTGGGCTCCAAAGTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGAT 1296
Db 1084 TCAGAGGCTGTGTGGGCTCCAAAGTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGAT 1143
QY 1297 CGAGGATGAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGGCTGG 1356
Db 1144 CGAGGATGAGTACAGCCACAGCATCATCAGAACCTCCTGTACCATGTGGTGGGCTGG 1203
QY 1357 ACCGACTGGAACCTTGGCCCTGAACCCCGAAGGAGGACCCCAATGGGTGCGTAACCTTGT 1416
Db 1204 ACCGACTGGAACCTTGGCCCTGAACCCCGAAGGAGGACCCCAATGGGTGCGTAACCTTGT 1263
QY 1417 GACAGTCCCATCATTTGTAGACATCACCAAGGACACCGTTTACAAACAGCCCATGTTCTAC 1476
Db 1264 GACAGTCCCATCATTTGTAGACATCACCAAGGACACCGTTTACAAACAGCCCATGTTCTAC 1323
QY 1477 CACCTTGGCCACTTCAGCAAGTTTCACTTCTGAGGGCTCCAGAGAGTGGGGCTGGTGGC 1536
Db 1324 CACCTTGGCCACTTCAGCAAGTTTCACTTCTGAGGGCTCCAGAGAGTGGGGCTGGTGGC 1383
QY 1537 AGTCAGAGAAACGACCTGGACGGAGTGGGACTGATGCATCCCGATGGCTCTGCTGTGTG 1596
Db 1384 AGTCAGAGAAACGACCTGGACGGAGTGGGACTGATGCATCCCGATGGCTCTGCTGTGTG 1443
```


QY 1597 GTCGTGCTAAACCGCTCCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGC 1656
|||||
Db 1444 GTCGTGCTAAACCGCTCCTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGC 1503
|||||
QY 1657 TTCCTGGAGACAATCTCACCTGGCTACTCCATTCACACCTACCTGTGGCGTCGCCAGTGA 1716
|||||
Db 1504 TTCCTGGAGACAATCTCACCTGGCTACTCCATTCACACCTACCTGTGGCGCGCTCAATGA 1563
|||||
RESULT 14
BC000349 1900 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, clone MGC:8614 IMAGE:2961326, mRNA, complete cds.
DEFINITION BC000349
ACCESSION BC000349
VERSION BC000349.1 GI:12653162
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1900)
Strausberg, R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 1 Row: m Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 496370.
Location/Qualifiers
1..1900

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:8614 IMAGE:2961326"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
377..1339
/codon_start=1
/product="Unknown (protein for MGC:8614)"
/protein_id="AAH00349.1"
/db_xref="GI:12653163"

CDS

/translation="MASCDFSIKTYVADTPDDFQLEHNSLPEEDPKLKIPLIHLALQ
LAQRPVSLASPWTSPRLTKTGAGNGKGLKGPRIYHOTWARYIVKFLDAYAEHK
LQFVAVTAENEPSAGLLSGYPFQCLGFTPEHQRIARDLGPPTLANGTHNVRLLMLD
DQRLPLPHWAKVVLTDPEAAKGLCGFQVLGAECARLLGSRDVAQPHHKKPPVPCGR
LDRLPEIIVDITKHTFYKOPMEFVHLGHFSKFIPEGSRVGLVASOKNDPDAVALMHPD
GSPVVVVLNRSSKDVPLTIKDPVGFLETISPYSIHTYLWRRQ"
471 a 524 c 506 g 399 t

BASE COUNT

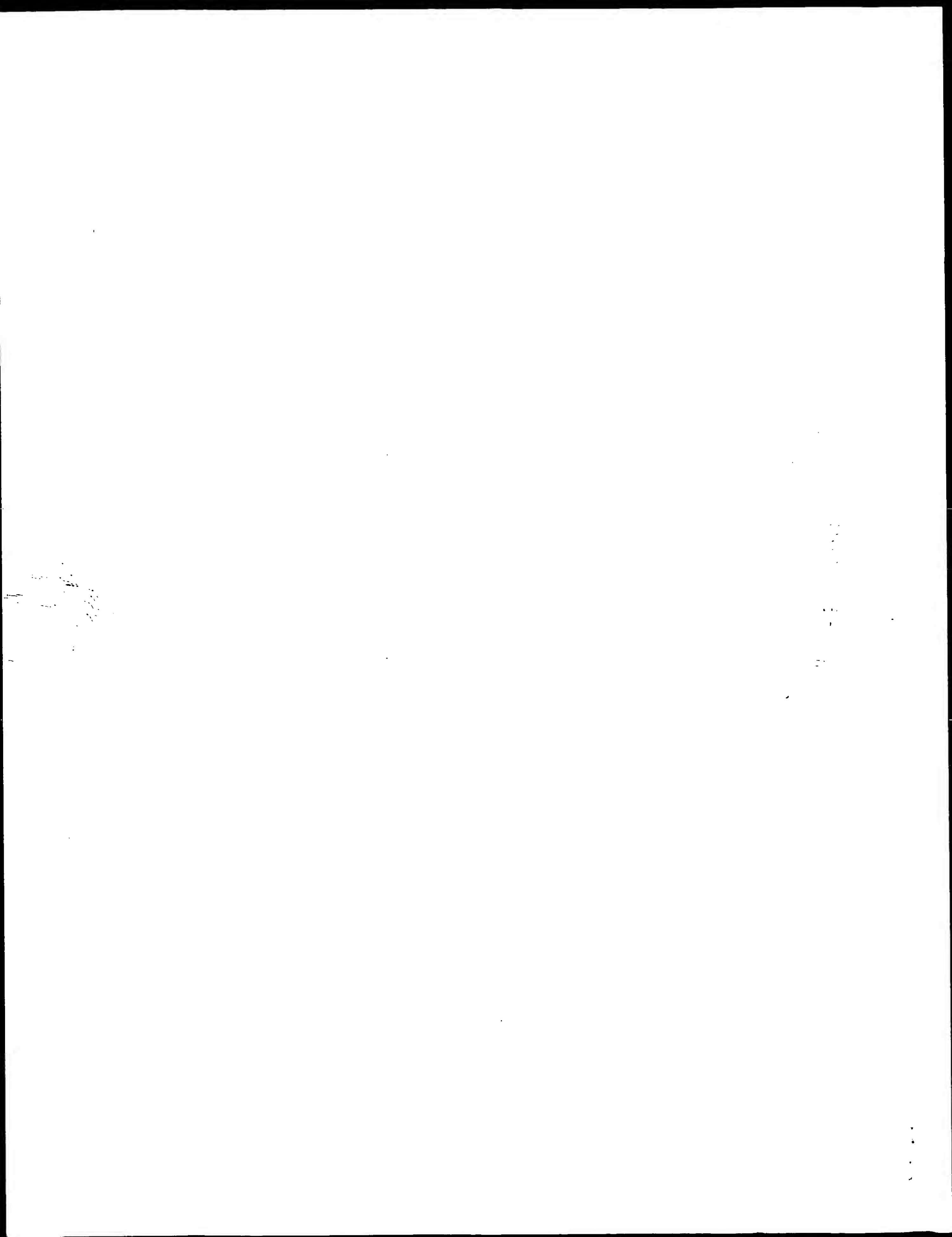
| | | | | | | |
|--------|------|---|--------------|--------------|------------|-------------------------|
| ORIGIN | | Query Match | 58.9%; | Score 1343; | DB 9; | Length 1900; |
| | | Best Local Similarity | 89.0%; | Pred. No. 0; | | |
| | | Matches 1543; | Conservative | 0; | Mismatches | 25; Indels 165; Gaps 2; |
| QY | 545 | ACTTCTCTGAAGAAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACT | 604 | | | |
| Db | 333 | ATTGCACCGGCACAGGAATCGGATATAACATCATCTGGGTACCCATGGCCAGCTGTGACT | 392 | | | |
| QY | 605 | TCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGATTCCAGTTGCACAACTTCA | 664 | | | |
| Db | 393 | TCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGATTCCAGTTGCACAACTTCA | 452 | | | |
| QY | 665 | GCCTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTACCCGAGCCCTGCAGTTGG | 724 | | | |
| Db | 453 | GCCTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTACCCGAGCCCTGCAGTTGG | 512 | | | |
| QY | 725 | CCCAGCGTCCCGTTTCACTCCTTCCAGCCCTGGACATCACCCACTTGGCTCAAGACCA | 784 | | | |
| Db | 513 | CCCAGCGTCCCGTTTCACTCCTTCCAGCCCTGGACATCACCCACTTGGCTCAAGACCA | 572 | | | |
| QY | 785 | ATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACCA | 844 | | | |
| Db | 573 | AGGAGCGGGGAATGGGAAGGGCCACTCAAGGGACAGCCCGGAGACATCTACCACCA | 632 | | | |
| QY | 845 | CCTGGGCCAGATACCTTTGTGAAGTTCTCTGGATGCCTATGCTGAGCACAAAGTTACAGTTCT | 904 | | | |
| Db | 633 | CCTGGGCCAGATACCTTTGTGAAGTTCTCTGGATGCCTATGCTGAGCACAAAGTTACAGTTCT | 692 | | | |
| QY | 905 | GGCAGTGACAGCTGAAATGAGCCTTCTCTGGGGCTGTTGAGTGGATACCCCTTCCAGT | 964 | | | |
| Db | 693 | GGCAGTGACAGCTGAAATGAGCCTTCTCTGGGGCTGTTGAGTGGATACCCCTTCCAGT | 752 | | | |
| QY | 965 | GCCTGGGCTTCAACCCCTGAACATCAGCGAGACTTCATTGCGCGTGACCTAGTCCCTACCC | 1024 | | | |
| Db | 753 | GCCTGGGCTTCAACCCCTGAACATCAGCGAGACTTCATTGCGCGTGACCTAGTCCCTACCC | 812 | | | |
| QY | 1025 | TCGCCAACAGTACTCACCAACAATGTCCGCTACTCATCTGGGATGACCAAGCTTGTCTGC | 1084 | | | |
| Db | 813 | TTGCCAACGCTACTCACCAACAATGTCCGCTACTCATCTGGGATGACCAAGCTTGTCTGC | 872 | | | |
| QY | 1085 | TGCCCCACTGGGCAAAAGGTGGTACTGACACACCCAGAACGAGCTAAATATGTTTCATGGCA | 1144 | | | |
| Db | 873 | TGCCCCACTGGGCAAAAGGTGGTACTGACACACCCAGAACGAGCTAA----- | 918 | | | |
| QY | 1145 | TTGCTGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCTTAGGGGAGACAC | 1204 | | | |
| Db | 919 | ----- | 918 | | | |
| QY | 1205 | ACCGCCTGTTCCTCAACACCATGCTCTTTGCTCCTCAGAGGCTGTGTGGCTCCCAAGTTCT | 1264 | | | |
| Db | 919 | -----AGGCTGTGTGGTTCCTCAAGTTCT | 942 | | | |
| QY | 1265 | GGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGCATCA | 1324 | | | |
| Db | 943 | GGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGCATCA | 1002 | | | |
| QY | 1325 | TCACGAACCTCCTGTACCATGTGGTGGTGGACCGGACTGGAACCTTGCCTGAACCCCG | 1384 | | | |
| Db | 1003 | TCACAAACCTCCTGTACCATGTGGTGGTGGACCGGACTGGA----- | 1045 | | | |
| QY | 1385 | AAGGAGACCCCAATTGGGTGCGTAACTTTGTGACAGTCCCATCATTTGTAGACATCACCA | 1444 | | | |
| Db | 1046 | -----CCCATCATTTGTAGACATCACCA | 1067 | | | |
| QY | 1445 | AGGACACGTTTACAAACAGCCCATGTTCTACCACCTTGGCCACTTCCAGCAAGTTCAATC | 1504 | | | |
| Db | 1068 | AGCACAGTTTACAAACAGCCCATGTTCTACCACCTTGGCCACTTCCAGCAAGTTCAATC | 1127 | | | |
| QY | 1505 | CTGAGGGTCCCAGAGAGTGGGGCTGGTTGCCAGTCAAGAACGACCTGGACGCGAGTGG | 1564 | | | |
| Db | 1128 | CTGAGGGTCCCAGAGAGTGGGGCTGGTTGCCAGTCAAGAACGACCTGGACGCGAGTGG | 1187 | | | |

QY 1565 CACTGATGATCCCGATGGCTCTGCTGTGTGGTGGTCTGCTAAACCGCTCCTCTAAGGATG 1624
|||||
Db 1188 CACTGATGATCCCGATGGCTCTGCTGTGTGGTGGTCTGCTAAACCGCTCCTCTAAGGATG 1247
|||||
QY 1625 TGCCTCTTACCATCAAGGATCCCTGCTGTGGCTTCCTGGAGACAAATCTCACCTGGCTACT 1684
|||||
Db 1248 TGCCTCTTACCATCAAGGATCCCTGCTGTGGCTTCCTGGAGACAAATCTCACCTGGCTACT 1307
|||||
QY 1685 CCATTACACCTTACCTGTGGCGTCCGCTGATGGAGCAGATCACTCAAGGAGGCACCTGGG 1744
|||||
Db 1308 CCATTACACCTTACCTGTGGCGTCCGCTGATGGAGCAGATCACTCAAGGAGGCACCTGGG 1367
|||||
QY 1745 CTCAGCCTGGGATTAAGGGACAGATCAGCTCACACGCTGTGTGTGACTAAAGAGGGC 1804
|||||
Db 1368 CTCAGCCTGGGATTAAGGGACAGATCAGCTCACACGCTGTGTGTGACTAAAGAGGGC 1427
|||||
QY 1805 ACAGCAGGGCCAGTGTGAGCTTACAGCGAGCTAAGCCCGAGGGCAATGGTTGGTGACT 1864
|||||
Db 1428 ACAACAGGGCCAGGCTGAGCTTACAGCGAGCTAAGCCCGAGGGCAATGGTTGGTGACT 1487
|||||
QY 1865 CACTTTCCCTCTAGGTGGTGGCGGCTGAGGCTGAGGCGCCCTAGAAAAAGATCAGTAAGCCC 1924
|||||
Db 1488 CACTTTCCCTCTAGGTGGTGGCGGCTGAGGCGCCCTAGAAAAAGATCAGTAAGCCC 1547
|||||
QY 1925 CAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTTTGGAA 1984
|||||
Db 1548 CAGTGTCCCCCAGCCCCCATGCTTATGTGAACATGCGCTGTGTGCTGCTTTGGAA 1607
|||||
QY 1985 ACTGGGCTGGGTCCAGGCTAGGTGAGTCACTGTCCGTACAAACACAGATCAGGGC 2044
|||||
Db 1608 ACTGGGCTGGGTCCAGGCTAGGTGAGTCACTGTCCGTACAAACACAGATCAGGGC 1667
|||||
QY 2045 TGAGGGTAAGGAAAGAGAGACTAGGAAAGCTGGGCCCCAAACCTGGAGACTGTTGTCT 2104
|||||
Db 1668 TGAGGGTAAGGAAAGAGAGACTAGGAAAGCTGGGCCCCAAACCTGGAGACTGTTGTCT 1727
|||||
QY 2105 TTCCTGGAGATGAGAACTGGGCCCCGTGGAGCAGCAGTGTGAGCTCAGGGCGGAGCCT 2164
|||||
Db 1728 TTCCTGGAGATGAGAACTGGGCCCCGTGGAGCAGCAGTGTGAGCTCAGGGCGGAGCCT 1787
|||||
QY 2165 TAAAGCAGCAGCGGTGTGCCAGGCAACCCAGATGATTCCTATGGCACCAGCCAGGAAAA 2224
|||||
Db 1788 TAAAGCAGCAGCGGTGTGCCAGGCAACCCAGATGATTCCTATGGCACCAGCCAGGAAAA 1847
|||||
QY 2225 ATGGCAGCTCTTAAAGGAGAAAAATGTTTGGCCCCAAAAAAGGAAAAA 2277
|||||
Db 1848 ATGGCAGCTCTTAAAGGAGAAAAATGTTTGGCCCCAAAAAAGGAAAAA 1900
|||||

RESULT 15
AK090514
LOCUS
DEFINITION Homo sapiens cDNA FLJ33195 fis, clone ADRGL2005961, highly similar
to GLUCOSYLKERAMIDASE PRECURSOR (EC 3.2.1.45).
ACCESSION AK090514
VERSION AK090514.1 GI:21748690
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens adrenal gland cDNA to mRNA, clone_lib:ADRGL2
clone:ADRGL2005961.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuo,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2711)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1. .2711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADRGL2005961"
/tissue_type="adrenal gland"
/clone_lib="ADRGL2"
/note="cloning vector: pME18SFL3"
BASE COUNT 622 a 789 c 688 g 612 t
ORIGIN

Query Match 56.5%; Score 1287; DB 9; Length 2711;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1517; Conservative 0; Mismatches 195; Indels 73; Gaps 4;
QY 547 TTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGACTTC 606
|||||
Db 927 TCCTTTTCTGTAGGAATCGGATATAACATCATCTGGGTACCCATGGCCAGCTGACTTC 986
|||||
QY 607 TCATCCGCACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGC 666
|||||
Db 987 TCATCCGCACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGC 1046
|||||
QY 667 CTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCGAGTTGGCC 726
|||||
Db 1047 CTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCGAGTTGGCC 1106
|||||
QY 727 CAGGCTCCCGTTTCACTCTTCCAGCCCTGGACATCACCCTGCTGAGACCAAT 786
|||||
Db 1107 CAGGCTCCCGTTTCACTCTTCCAGCCCTGGACATCACCCTGCTGAGACCAAG 1166
|||||
QY 787 GGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGGAGACATCTACACAGACC 846
|||||
Db 1167 GGAGCGGGGAATGGGAAGGGGCACTCAAGGGACAGCCCGGAGACATCTACACAGACC 1226
|||||
QY 847 TGGCCAGAGATACTTTGTGAAGTTCTTGGATGCTTATGCTGAGCACAAAGTTACAGTTCTGG 906
|||||
Db 1227 TGGCCAGAGATACTTTGTGAAGTTCTTGGATGCTTATGCTGAGCACAAAGTTACAGTTCTGG 1286
|||||
QY 907 GCAGTGACAGCTGAAAATAGCCTTCTGTGGGCTGTGAGTGGATACCCCTTCCAGTGC 966
|||||
Db 1287 GCAGTGACAGCTGAAAATAGCCTTCTGTGGGCTGTGAGTGGATACCCCTTCCAGTGC 1346
|||||
QY 967 CTGGCTTACCCCTGAACATCAGCGAGACTTCAATGCGCGTGACCTAGCTTACCTTACCCCTC 1026
|||||
Db 1347 CTGGCTTACCCCTGAACATCAGCGAGACTTCAATGCGCGTGACCTAGCTTACCTTACCCCTT 1406
|||||
QY 1027 GCCAACAGTACTCACCACATGTCCGCTTACTCATGCTGGATGACCAACGCTTGTGCTG 1086
|||||
Db 1407 GCCAACGCTACTCACCACATGTCCGCTTACTCATGCTGGATGACCAACGCTTGTGCTG 1466
|||||
QY 1087 CCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGAGCAAGTAAATATGTTTACGTCATT 1146
|||||
Db 1467 CCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGAGCAAGTAAAGGCTGTGTGGGTTTC 1526
|||||
QY 1147 GCTGTAC-----ATTGGTACCTGGACTTCTGG-----CTCCAGGCCA 1183
|||||
Db 1527 CAAGTCTGGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGGCCA 1586
|||||



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:17:27 ; Search time 31 Seconds
(without alignments)
508.732 Million cell updates/sec

Title: US-10-024-197-25
Perfect score: 2865
Sequence: 1 MEFSSPSRECPKPLSRVSI.....GFLETISPGYSIHTYLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCrUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 2864 | 100.0 | 546 | 2 | US-08-713-928B-4 |
| 2 | 2665 | 93.0 | 498 | 1 | US-08-080-855-1 |
| 3 | 105.5 | 3.7 | 430 | 2 | US-08-924-440-2 |
| 4 | 97 | 3.4 | 1711 | 2 | US-08-342-930-2 |
| 5 | 94.5 | 3.3 | 180 | 2 | US-08-924-759-24 |
| 6 | 94.5 | 3.3 | 180 | 3 | US-09-248-335-24 |
| 7 | 93 | 3.2 | 15281 | 2 | US-08-471-119A-2 |
| 8 | 92 | 3.2 | 716 | 4 | US-08-971-188-5 |
| 9 | 92 | 3.2 | 716 | 4 | US-08-971-188-11 |
| 10 | 90.5 | 3.2 | 437 | 4 | US-09-801-052-4 |
| 11 | 90.5 | 3.2 | 488 | 4 | US-09-801-052-5 |
| 12 | 88.5 | 3.1 | 4302 | 3 | US-08-658-136-5 |
| 13 | 88.5 | 3.1 | 4302 | 4 | US-09-052-469-8 |
| 14 | 88.5 | 3.1 | 4303 | 2 | US-08-460-751-2 |
| 15 | 88.5 | 3.1 | 4339 | 4 | US-09-052-469-6 |
| 16 | 88 | 3.1 | 556 | 2 | US-09-016-000-4 |
| 17 | 86.5 | 3.0 | 2629 | 2 | US-08-751-189-4 |
| 18 | 86.5 | 3.0 | 2629 | 2 | US-09-060-836-4 |
| 19 | 86.5 | 3.0 | 2629 | 4 | US-09-184-445-4 |
| 20 | 86 | 3.0 | 675 | 4 | US-08-485-355B-52 |
| 21 | 85.5 | 3.0 | 465 | 4 | US-09-801-052-2 |
| 22 | 85.5 | 3.0 | 591 | 3 | US-08-965-903B-2 |
| 23 | 85 | 3.0 | 501 | 3 | US-08-906-791-2 |
| 24 | 84.5 | 2.9 | 728 | 4 | US-08-915-337-2 |
| 25 | 84 | 2.9 | 545 | 4 | US-09-066-046-39 |
| 26 | 83.5 | 2.9 | 274 | 1 | US-08-318-947A-20 |
| 27 | 83.5 | 2.9 | 274 | 2 | US-08-795-303-20 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 28 | 83 | 2.9 | 690 | 4 | US-09-228-986-69 | Sequence 69, Appl |
| 29 | 82 | 2.9 | 1323 | 1 | US-08-026-138E-4 | Sequence 4, Appli |
| 30 | 81.5 | 2.8 | 737 | 4 | US-08-971-188-4 | Sequence 4, Appli |
| 31 | 81.5 | 2.8 | 852 | 2 | US-08-408-519-5 | Sequence 5, Appli |
| 32 | 81.5 | 2.8 | 852 | 5 | PCT-US95-03552-5 | Sequence 5, Appli |
| 33 | 81.5 | 2.8 | 953 | 4 | US-09-245-281-43 | Sequence 43, Appl |
| 34 | 81.5 | 2.8 | 953 | 4 | US-09-207-359B-43 | Sequence 4, Appli |
| 35 | 81 | 2.8 | 597 | 1 | US-08-399-646-4 | Sequence 4, Appli |
| 36 | 81 | 2.8 | 597 | 1 | US-08-607-321-4 | Sequence 4, Appli |
| 37 | 81 | 2.8 | 597 | 2 | US-08-961-240-4 | Sequence 4, Appli |
| 38 | 81 | 2.8 | 597 | 2 | US-08-605-501-4 | Sequence 4, Appli |
| 39 | 81 | 2.8 | 598 | 1 | US-08-399-646-14 | Sequence 14, Appl |
| 40 | 81 | 2.8 | 598 | 1 | US-08-607-321-14 | Sequence 14, Appl |
| 41 | 81 | 2.8 | 598 | 2 | US-08-961-240-14 | Sequence 14, Appl |
| 42 | 81 | 2.8 | 598 | 2 | US-08-605-501-14 | Sequence 14, Appl |
| 43 | 80.5 | 2.8 | 212 | 2 | US-08-924-759-14 | Sequence 14, Appl |
| 44 | 80.5 | 2.8 | 212 | 3 | US-09-248-335-14 | Sequence 14, Appl |
| 45 | 80.5 | 2.8 | 580 | 4 | US-09-188-930-307 | Sequence 307, App |

ALIGNMENTS

RESULT 1
US-08-713-928B-4
; Sequence 4, Application US/08713928B
; Patent No. 5929304
; GENERAL INFORMATION:
; APPLICANT: RADIN, DAVID N.
; APPLICANT: CRAMER, CAROLE L.
; APPLICANT: OISHI, KAREN K.
; APPLICANT: WEISSENBORN, DEBORAH L.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,928B
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/003,737
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7956-0011-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-713-928B-4

Query Match 100.0%; Score 2864; DB 2; Length 546;
Best Local Similarity 99.8%; Pred. No. 2.1e-304;

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICANT: MAURO, LAURA J.
FILING DATE: August 27, 1997
PRIORITY APPLICATION DATA:
APPLICANT: DIXON, JACK E.

ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 846-7555
TELEFAX: (650) 845-6504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-924-440-2

Query Match 3.7%; Score 105.5; DB 2; Length 430;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 98; Conservative 53; Mismatches 150; Indels 169; Gaps 30;

QY 34 VSWASGARPCIPKSGY-----SSVVCVCNATYCD-----SFPDPTFPAL 73
Db 66 VSWNSTIPPGTASGFIASGSEPHCTINGAPCDGSEPGGPGGPGTSPDPGTPGT 125
QY 74 GT-PSRYESTR-SGRMELSMG-PIQANHTGTGLLLTLQPEQKFKVKGFGGAMTDAAL 130
Db 126 GTPVRYGKVQVCGQLCDEHGNPVQLRGMST-----HGIQWFDHCLTD-SSL 172
QY 131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCD-----FSIRT-----YTYAD-----T 177
Db 173 DALAYDWKADIIRLSMYIQEDGYETN-----PRGFTDRMHQLIDMATARGLYVIVDWHILT 228
QY 178 PDDEFQHNFLPEEDTKLKIPLIHRALQLAQRPVSLASPWTSPWLKTNAGVNGKSLK 237
Db 229 PGD-----PHYNLDRAKTFE-----AEIAQRHAS-----KTNVLY-----EIA 261
QY 238 GQPDYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSG-YPFQCLGFTPEHQR 296
Db 262 NEPNGV---SWAS-----IKSYAEV---PVIRQDPDSVVIIVTRGWSL-----303
QY 297 DFLARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIHVHLYL--- 353
Db 304 --VSEGSGP-----AEIAANPVNASNIMY--AFHFYAASH 334
QY 354 --DFLAPAKATLGETHRLFPNTMLFASACVGSKEFWQSVRLGSDWRGMQYSHSITNLL 411
Db 335 RDNLYL---NALREASELFP---VFVTE---FGTETY---TGDGANDFQMAADRYIDLMA 380
QY 412 YHVVGWTDWNL-----LNP-----EGGP-----NWRNEVDS 439
Db 381 ERKIGWTKWNSDDFRSGAVFQPGTCASGGPWSGSSSLKASGQWVRSLQS 430

RESULT 4
US-08-342-930-2
Sequence 2, Application US/08342930
Patent No. 5821084
GENERAL INFORMATION:
APPLICANT: OLMSTED, ELIZABETH A.
APPLICANT: MAURO, LAURA J.
APPLICANT: DAVIS, ALAN R.
APPLICANT: DIXON, JACK E.
TITLE OF INVENTION: OSTEOBLAST-TESTICULAR PROTEIN TYROSINE

TITLE OF INVENTION: PHOSPHATASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,930
FILING DATE: 21-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-20975.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1711 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-342-930-2

Query Match 3.4%; Score 97; DB 2; Length 1711;
Best Local Similarity 19.6%; Pred. No. 1.3;
Matches 107; Conservative 60; Mismatches 158; Indels 220; Gaps 31;

QY 131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTYAD--TPDDFQLHNFSL 188
Db 519 DLVDLGPDTLSTLTKSL-----VP-GSC-YIVSAWAWAGNLSDDSKIHSCSTR 564
QY 189 PEEDTKLKIPLIHRALQLAQRPVSLASPWTSPWLKTNAGVNGK-----234
Db 565 PAPPTNL-----SLGFAHQPAALKAS-WYHPP-----GGRDAFHLRLRLRPLTL 608
QY 235 -SLKGOPGDIYHQTWARYFVKFLDAYAEHKLQ--FWAVTAENEPSAGLLSGYFPQCLGF 290
Db 609 ESEKVLPREAQNFWSAQ-----LTAGCEFQVQLSTLWG--SERSSA-----NATGW 653
QY 291 TPEHQDFIARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKV-----VLTDPAAK 342
Db 654 TPTS-----APTLVNVT-----DAPTQLQVSWAHVPGGSRRYQVTLYQESTR 696
QY 343 YVHGI-----AVHWYLDFLAPAKATLGETHRLFPNTMLFAS- 379
Db 697 TATSIMGPKEDGTSFLGLTPGTYKYKVEVISWAGPLYTAAANVSATVPLIPNELLVSMQA 756
QY 380 -----AC-----VGSKFWEQSVRLGS--WDRGMQYSHSI-----406
Db 757 GSAVVNLAWPSGPLGQGACHAQLSDAHLSWEQPLKLGQELFMLRLDLPHTTISMSVRCR 816
QY 407 -----ITNLLYH-----VVGWT-----DWNL-----LNPEGPN 431
Db 817 AGPLQASTHLVLSVBPVDPVEDVLCHPEATYALNWTMPAGDVCLVSVVVERLVPGSGTH 876
QY 432 W---VNRNVDSPIIIVDITKDTFYKQPMFYHLGHFSKFIPEGSRQVGLVASQKNDLDAVAL 488
Db 877 FVFQVNTSGDALLPLNLMPTTSYRLSLTV-LGRNSRW-----SRAVSLVCS-----TSAEA 926
QY 489 MHP-----DGSVVVVLNRSSKDVPLTIKDPV--GFLETIS-----PGYSI 528
Db 927 WHPPELAEPQVELGTGMGVTVMRGMFGKD-----DGIQWYGIATINMTLAQPSREA 980

QY 529 HTYLW 533
Db 981 INYTW 985

RESULT 5

US-08-924-759-24
; Sequence 24, Application US/08924759
; Patent No. 5962229

GENERAL INFORMATION:

APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/924,759

FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1128
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: MAIZE
IMMEDIATE SOURCE:
LIBRARY: CS1.PK0059.E2

US-08-924-759-24

Query Match 3.3%; Score 94.5; DB 2; Length 180;

Best Local Similarity 25.4%; Pred. No. 0.047;

Matches 36; Conservative 22; Mismatches 55; Indels 29; Gaps 7;

QY 143 LLKSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHR 202

Db 9 LLGSWYSP-----YVIRAKVA---LGLKGLSYEFVEEDLS-----RKSDLLKLNPNVHR 54

QY 203 ALQL---AQRPV---SLLASPTWTSPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLD 257

Db 55 KVPVLVHGGRPVCSLSVILQYVDETWTAGTGTPL-----LPADAYDRAMARFWAAAYVD 106

QY 258 --AYAETHKQFWAVTAENEPSA 277

Db 107 DKFYKEWNRLFWSTTAEKAAEA 128

RESULT 6

US-09-248-335-24

; Sequence 24, Application US/09248335
; Patent No. 6096504

GENERAL INFORMATION:

APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEF, DANIEL
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1128-A
CURRENT APPLICATION NUMBER: US/09/248,335
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 08/924,759
EARLIER FILING DATE: 1997-September-05
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 24
LENGTH: 180
TYPE: PRT
ORGANISM: maize
US-09-248-335-24

Query Match 3.3%; Score 94.5; DB 3; Length 180;

Best Local Similarity 25.4%; Pred. No. 0.047;

Matches 36; Conservative 22; Mismatches 55; Indels 29; Gaps 7;

QY 143 LLKSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHR 202

Db 9 LLGSWYSP-----YVIRAKVA---LGLKGLSYEFVEEDLS-----RKSDLLKLNPNVHR 54

QY 203 ALQL---AQRPV---SLLASPTWTSPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLD 257

Db 55 KVPVLVHGGRPVCSLSVILQYVDETWTAGTGTPL-----LPADAYDRAMARFWAAAYVD 106

QY 258 --AYAETHKQFWAVTAENEPSA 277

Db 107 DKFYKEWNRLFWSTTAEKAAEA 128

RESULT 7

US-08-471-119A-2

; Sequence 2, Application US/08471119A

; Patent No. 5827706

GENERAL INFORMATION:

APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoengendorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5827706gartis Corporation

STREET: 59 Route 10

CITY: East Hanover

STATE: New Jersey

COUNTRY: USA

ZIP: 07936

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,119A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kassenoff, Melvyn

REGISTRATION NUMBER: 26,389

REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 503 8474

TELEFAX: 201 503 8807

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 15281 amino acids

TYPE: amino acid

STRANDEDNESS: single


```

; Sequence 8, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-052-469-8

```

```

Query Match      3.1%; Score 88.5; DB 4; Length 4302;
Best Local Similarity 22.0%; Pred. No. 57;
Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;

QY 269 VTAENEPSAG-----LLSGYFFQCLG-----FTPEHQDFTA----- 300
Db 1611 VTAENEVGSQDSIFVYVQLIEG--LQVVGGRYFPTNHTVQLQAVVRDGTNVSYSWTA 1668
QY 301 -RDLGPTLANS-----THNVRLLMLDDQRLLLPHWAKVVLTDPEAAKYVHGI 347
Db 1669 WRDRGPALAGSGKGFSLTVLEAGTYH-----VQLRATNMLGSAWADCTMDFE----- 1716
QY 348 AVHWYLDLFLAPAKATLGETHRLFPNTMLFASEACVGSK--FWEQSVRLG-SWDRGMQY-S 403
Db 1717 PVGWLMTASPNPAAV-----NTSVTLAELAGSGGVVYTWLSLEGLSWETSEPTT 1768
QY 404 HSIITNLLYHVVGWTDWNLALNPEGPNWVRNFVDSPIIVDITKDTFYKQPM----- 455
Db 1769 HSFPTPGL-HLVTMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1810
QY 456 -----FYHLGHFSKF-----IPEGQRVGLVASQKNLDLDAVALMHPDGA 495

```

```

Db 1811 EPGGSFVAAGSSVPFWGQLATGTNVSWCWAIVPGSSKRG-----PHVTMVFPDAGT 1861
QY 496 VVVVLNRS-----SKDVPLTIKDPVAGFL-----ETISPGYSIH 529
Db 1862 FSIRLNASNAVSWVSATYNLTAEPIVGLVWLWASSKVAVPAGQLVH 1906

RESULT 14
US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-460-751-2

Query Match      3.1%; Score 88.5; DB 2; Length 4303;
Best Local Similarity 22.0%; Pred. No. 57;
Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;

QY 269 VTAENEPSAG-----LLSGYFFQCLG-----FTPEHQDFTA----- 300
Db 1611 VTAENEVGSQDSIFVYVQLIEG--LQVVGGRYFPTNHTVQLQAVVRDGTNVSYSWTA 1668
QY 301 -RDLGPTLANS-----THNVRLLMLDDQRLLLPHWAKVVLTDPEAAKYVHGI 347
Db 1669 WRDRGPALAGSGKGFSLTVLEAGTYH-----VQLRATNMLGSAWADCTMDFE----- 1716
QY 348 AVHWYLDLFLAPAKATLGETHRLFPNTMLFASEACVGSK--FWEQSVRLG-SWDRGMQY-S 403
Db 1717 PVGWLMTASPNPAAV-----NTSVTLAELAGSGGVVYTWLSLEGLSWETSEPTT 1768
QY 404 HSIITNLLYHVVGWTDWNLALNPEGPNWVRNFVDSPIIVDITKDTFYKQPM----- 455
Db 1769 HSFPTPGL-HLVTMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1810

```

QY 456 -----FYHLGHFSKF-----IPEGQRVGLVASQKNDLDAVALMHPDGS 495
Db 1811 EPGGSFVAAGSSVPFWGQLATGTNVSWCWAVPGSSKRG-----PHVTMVFPDAGT 1861
QY 496 VVVVLNRS-----SKDVPLTIKDPVAVGFL-----ETISPGYSIH 529
Db 1862 FSIRLNASNAVSWVSATYNLTAEPIVGLVLWASSKVVVAPGQLVH 1906

RESULT 15
US-09-052-469-6
; Sequence 6, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-052-469-6

Query Match 3.18; Score 88.5; DB 4; Length 4339;
Best Local Similarity 22.0%; Pred. No. 58;
Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;
QY 269 VTAENEPSAG-----LLSGYFPQCLG-----FTPEHQDFIA----- 300
Db 1568 VTAENEVGSQDSIFVYVLQIEG--LQVVGGRYFPTNHTVOLQAVVRDGTNVSYSWTA 1625
QY 301 -RDLGPTLANS-----THNVRLLMLDDQRLLLPHWAKVWLTDPEAAKYVHGI 347

Db 1626 WRDRGPALAGSGKGFSLTVLEAGTYH-----VQLRATNMLGSAWADCTMDFE----- 1673
QY 348 AVHWYLDFLAPAKATLGETHRLFPNTMLFASEACVGSK--FWEQSVRLG-SWDRGMQY-S 403
Db 1674 PVGWLMTASPNPAAV-----NTSVTLSAELAGGSGVVYTWLSLEGLSWETSEPFPT 1725
QY 404 HSIITNLLYHVVGWTDWNLALNPEGGNWVRNFVDSPIIVDITKDTFYKQPM----- 455
Db 1726 HSFPPTPGL-HLVTMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1767
QY 456 -----FYHLGHFSKF-----IPEGQRVGLVASQKNDLDAVALMHPDGS 495
Db 1768 EPGGSFVAAGSSVPFWGQLATGTNVSWCWAVPGSSKRG-----PHVTMVFPDAGT 1818
QY 496 VVVVLNRS-----SKDVPLTIKDPVAVGFL-----ETISPGYSIH 529
Db 1819 FSIRLNASNAVSWVSATYNLTAEPIVGLVLWASSKVVVAPGQLVH 1863

Search completed: February 20, 2003, 11:21:03
Job time : 47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:07:51 ; Search time 39 Seconds
(without alignments)
1831.342 Million cell updates/sec

Title: US-10-024-197-25
Perfect score: 2865
Sequence: 1 MEFFSPSRECKPLSRVSI.....GFLETISPGYSIHTYLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 2861 | 99.9 | 546 | AAW18237 | Human glucocerebro |
| 2 | 2854 | 99.6 | 536 | AAW18237 | Protein encoded by |
| 3 | 2854 | 99.6 | 536 | AAW18237 | protein encoded by |
| 4 | 2832 | 98.8 | 536 | AAE02446 | Human glucocerebro |
| 5 | 2762 | 96.4 | 516 | AAE02446 | Human lysosomal gl |
| 6 | 2762 | 96.4 | 516 | AAE02446 | Human lysosomal gl |
| 7 | 2762 | 96.4 | 516 | AAE02446 | Amino acid sequenc |
| 8 | 2757 | 96.2 | 516 | AAE02446 | Human glucocerebro |
| 9 | 2757 | 96.2 | 516 | AAE02446 | Amino acid sequenc |
| 10 | 2736 | 95.5 | 516 | AAE02446 | Recombinant human |

| | | | | | | |
|----|------|------|-----|----|----------|--------------------|
| 11 | 2675 | 93.4 | 592 | 22 | AAU05698 | Human glucocerebro |
| 12 | 2665 | 93.0 | 498 | 17 | AAW07885 | Remodelled, recomb |
| 13 | 2664 | 93.0 | 497 | 22 | AAU05696 | Human Mature gluco |
| 14 | 2664 | 93.0 | 497 | 23 | AAU86020 | Human glucocerebro |
| 15 | 2659 | 92.8 | 497 | 22 | AAU05668 | Human glucocerebro |
| 16 | 2659 | 92.8 | 497 | 22 | AAU05680 | Human glucocerebro |
| 17 | 2659 | 92.8 | 497 | 22 | AAU05684 | Human glucocerebro |
| 18 | 2658 | 92.8 | 497 | 22 | AAU05699 | Human glucocerebro |
| 19 | 2656 | 92.7 | 497 | 22 | AAU05667 | Human glucocerebro |
| 20 | 2656 | 92.7 | 497 | 22 | AAU05695 | Human glucocerebro |
| 21 | 2655 | 92.7 | 497 | 22 | AAU05676 | Human glucocerebro |
| 22 | 2655 | 92.7 | 497 | 22 | AAU05678 | Human glucocerebro |
| 23 | 2655 | 92.7 | 497 | 22 | AAU05681 | Human glucocerebro |
| 24 | 2655 | 92.7 | 497 | 22 | AAU05685 | Human glucocerebro |
| 25 | 2655 | 92.7 | 497 | 22 | AAU05692 | Human glucocerebro |
| 26 | 2654 | 92.6 | 497 | 22 | AAU05654 | Human glucocerebro |
| 27 | 2654 | 92.6 | 497 | 22 | AAU05656 | Human glucocerebro |
| 28 | 2654 | 92.6 | 497 | 22 | AAU05662 | Human glucocerebro |
| 29 | 2654 | 92.6 | 497 | 22 | AAU05666 | Human glucocerebro |
| 30 | 2654 | 92.6 | 497 | 22 | AAU05670 | Human glucocerebro |
| 31 | 2654 | 92.6 | 497 | 22 | AAU05671 | Human glucocerebro |
| 32 | 2654 | 92.6 | 497 | 22 | AAU05675 | Human glucocerebro |
| 33 | 2654 | 92.6 | 497 | 22 | AAU05688 | Human glucocerebro |
| 34 | 2653 | 92.6 | 497 | 22 | AAU05653 | Human glucocerebro |
| 35 | 2653 | 92.6 | 497 | 22 | AAU05655 | Human glucocerebro |
| 36 | 2653 | 92.6 | 497 | 22 | AAU05661 | Human glucocerebro |
| 37 | 2653 | 92.6 | 497 | 22 | AAU05665 | Human glucocerebro |
| 38 | 2653 | 92.6 | 497 | 22 | AAU05669 | Human glucocerebro |
| 39 | 2653 | 92.6 | 497 | 22 | AAU05672 | Human glucocerebro |
| 40 | 2653 | 92.6 | 497 | 22 | AAU05673 | Human glucocerebro |
| 41 | 2653 | 92.6 | 497 | 22 | AAU05674 | Human glucocerebro |
| 42 | 2653 | 92.6 | 497 | 22 | AAU05677 | Human glucocerebro |
| 43 | 2653 | 92.6 | 497 | 22 | AAU05679 | Human glucocerebro |
| 44 | 2653 | 92.6 | 497 | 22 | AAU05689 | Human glucocerebro |
| 45 | 2653 | 92.6 | 497 | 22 | AAU05691 | Human glucocerebro |

ALIGNMENTS

RESULT 1
AAW18237
ID AAW18237 standard; Protein; 546 AA.
XX
AC AAW18237;
XX
DT 01-OCT-1997 (first entry)
XX
DE Human glucocerebrosidase-FLAG fusion.
XX
KW Lysosomal enzyme; lysosome; transgenic plant; glucocerebrosidase;
XX enzyme replacement therapy; Gaucher disease.
XX
OS Chimaeric Homo sapiens;
OS Chimaeric synthetic.
XX
FH Key Location/Qualifiers
FT Protein 1..536
FT Peptide /label= hCG
FT /label= FLAG
FT Misc-difference 89
FT /note= "encoded by GGG"
XX
PN WO9710353-A1.
XX
PD 20-MAR-1997.
XX
PF 13-SEP-1996; 96WO-US14730.
XX
PR 14-SEP-1995; 95US-0003737.
XX
PA (CROP-) CROPTech DEV CORP.

PA (VIRG) VIRGINIA TECH INTELLECTUAL PTY INC.
XX Cramer CL, Oishi KK, Radin DN, Weissenborn DL;
PI WPI; 1997-202248/18.
XX N-PSDB; AAT71753.
XX Production of enzymatically active (modified) lysosomal enzyme in
PT transgenic plants - useful in treatment of lysosomal storage
PT disorders
XX Example 1; Page 62-64; 111pp; English.
PS A novel fusion protein (AAW18237) comprises human glucocerebrosidase
XX (hGC) and a C-terminal FLAG epitope (see also AAW18236). A DNA
CC construct (see also AAT71753) encoding the fusion is placed under
CC control of the inducible MEGA promoter (see also AAT71752) in a
CC plant expression system to provide production of recombinant hGC
CC in transgenic (esp. tobacco) plants. The plant expression system
CC provides for post-translational modification and processing to
CC produce enzymatically active hGC, which can be harvested for use in
CC Gaucher disease enzyme replacement therapy.
XX Sequence 546 AA;
SQ Query Match 99.9%; Score 2861; DB 18; Length 546;
Best Local Similarity 99.6%; Pred. No. 2.1e-290;
Matches 534; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEFSPPSREECPLSRVSIMAGSLTGLLLQAVSWASGAPCIPKSFYSSVVCVNCAT 60
Db 1 MEFSPPSREECPLSRVSIMAGSLTGLLLQAVSWASGAPCIPKSFYSSVVCVNCAT 60
QY 61 YCDSFDPTTFFALGTFSTRSRRMELSMGPIQANHTGTGLLTLQPEQKFKVKGF 120
Db 61 YCDSFDPTTFFALGTFSTRSRRMELSMGPIQANHTGTGLLTLQPEQKFKVKGF 120
QY 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180
Db 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180
QY 181 FQLHNFSLPEEDTKLKIPLIHRALQAPVSLASPWTSPWLKTNAGVNGKSLKGQP 240
Db 181 FQLHNFSLPEEDTKLKIPLIHRALQAPVSLASPWTSPWLKTNAGVNGKSLKGQP 240
QY 241 GDYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHQDFIA 300
Db 241 GDYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHQDFIA 300
QY 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
Db 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
QY 361 ATLGETHRLFPNTMLFASEACVSGKFWESQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
Db 361 ATLGETHRLFPNTMLFASEACVSGKFWESQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
QY 421 NLALNPEGGPNNVRNFVDSPIIVDTRDTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQK 480
Db 421 NLALNPEGGPNNVRNFVDSPIIVDTRDTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQK 480
QY 481 NDLDVALMHPDGSAAVVVNLNRSSKDVPDLTIKDPVGFLETISPGYSIHTYLWRRQ 536
Db 481 NDLDVALMHPDGSAAVVVNLNRSSKDVPDLTIKDPVGFLETISPGYSIHTYLWRRQ 536
RESULT 2
AAB61495
ID AAB61495 standard; Protein; 536 AA.
XX AAB61495;
AC AAB61495;
XX 04-APR-2001 (first entry)

XX Protein encoded by unmodified Gcc DNA.
DE Glucocerebrosidase; Gcc; splice; gene therapy; enzyme therapy;
XX Gaucher.
KW Unidentified.
XX CA2272055-A1.
XX 02-DEC-2000.
XX 02-JUN-1999; 99CA-2272055.
XX 02-JUN-1999; 99CA-2272055.
PA (HSCR-) HSC RES & DEV LP.
XX Mahuran DJ, Callahan JW, Clarke JTR;
XX WPI; 2001-160192/17.
XX New glucocerebrosidase (Gcc) DNA molecule modified in at least one
PT nucleotide, useful in gene therapy or enzyme replacement therapy,
PT particularly in treating Gaucher disease -
XX Disclosure; Fig 4; 37pp; English.
PS The present invention relates to glucocerebrosidase (Gcc) DNA
XX with a modification in at least one nucleotide that disrupts a
CC splicing consensus sequence and prevents splicing of mRNA produced
CC from the DNA molecule, while preserving the ability of the DNA to
CC express active Gcc. The glucocerebrosidase (Gcc) DNA molecule and
CC the vector comprising the DNA are useful in gene therapy or enzyme
CC replacement therapy, particularly in treating Gaucher disease. The
CC cells transfected or transduced by the vectors are useful as research
CC tools.
XX Sequence 536 AA;
SQ Query Match 99.6%; Score 2854; DB 22; Length 536;
Best Local Similarity 99.6%; Pred. No. 1.1e-289;
Matches 534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 NEFSPPSREECPLSRVSIMAGSLTGLLLQAVSWASGAPCIPKSFYSSVVCVNCAT 60
Db 1 NEFSPPSREECPLSRVSIMAGSLTGLLLQAVSWASGAPCIPKSFYSSVVCVNCAT 60
QY 61 YCDSFDPTTFFALGTFSTRSRRMELSMGPIQANHTGTGLLTLQPEQKFKVKGF 120
Db 61 YCDSFDPTTFFALGTFSTRSRRMELSMGPIQANHTGTGLLTLQPEQKFKVKGF 120
QY 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180
Db 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180
QY 181 FQLHNFSLPEEDTKLKIPLIHRALQAPVSLASPWTSPWLKTNAGVNGKSLKGQP 240
Db 181 FQLHNFSLPEEDTKLKIPLIHRALQAPVSLASPWTSPWLKTNAGVNGKSLKGQP 240
QY 241 GDYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHQDFIA 300
Db 241 GDYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHQDFIA 300
QY 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
Db 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
QY 361 ATLGETHRLFPNTMLFASEACVSGKFWESQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
Db 361 ATLGETHRLFPNTMLFASEACVSGKFWESQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
QY 421 NLALNPEGGPNNVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQK 480

Db 421 NLALNPEGPNWVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPESQVRGLVASQK 480
QY 481 NDLDVALMHDPDGSVAVVVLNRSSKDVPLTIKDPVAGFLETISPGYSIHITLWRRQ 536
Db 481 NDLDVALMHDPDGSVAVVVLNRSSKDVPLTIKDPVAGFLETISPGYSIHITLWRRQ 536
RESULT 3
AAB61496
ID AAB61496 standard; Protein; 536 AA.
XX
AC AAB61496;
XX
DT 04-APR-2001 (first entry)
XX
DE protein encoded by modified Gcc DNA.
XX
KW Glucocerebrosidase; Gcc; splice; gene therapy; enzyme therapy;
KW Gaucher.
XX
OS Unidentified.
XX
PN CA2272055-A1.
XX
PD 02-DEC-2000.
XX
PF 02-JUN-1999; 99CA-2272055.
XX
PR 02-JUN-1999; 99CA-2272055.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Mahuran DJ, Callahan JW, Clarke JTR;
XX
DR WPI; 2001-160192/17.
XX
PT New glucocerebrosidase (Gcc) DNA molecule modified in at least one
PT nucleotide, useful in gene therapy or enzyme replacement therapy,
PT particularly in treating Gaucher disease -
XX
PS Disclosure; Fig 4; 37pp; English.
XX
CC The present invention relates to glucocerebrosidase (Gcc) DNA
CC with a modification in at least one nucleotide that disrupts a
CC splicing consensus sequence and prevents splicing of mRNA produced
CC from the DNA molecule, while preserving the ability of the DNA to
CC express active Gcc. The glucocerebrosidase (Gcc) DNA molecule and
CC the vector comprising the DNA are useful in gene therapy or enzyme
CC replacement therapy, particularly in treating Gaucher disease. The
CC cells transfected or transduced by the vectors are useful as research
CC tools.
XX
SQ Sequence 536 AA;
Query Match 99.6%; Score 2854; DB 22; Length 536;
Best Local Similarity 99.6%; Pred. No. 1.1e-289;
Matches 534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MEFSSPSREECPKPLSRVSIMAGSLTGLLLQAVSWASGARPICPKSFYSSVVCVNCAT 60
Db 1 MEFSSPSREECPKPLSRVSIMAGSLTGLLLQAVSWASGARPICPKSFYSSVVCVNCAT 60
QY 61 YCDSFDPPPTFPALGTFSTRYESTRSGRRMELSMGPIQANHTGTGLLLTLQPEKFKQVKGF 120
Db 61 YCDSFDPPPTFPALGTFSTRYESTRSGRRMELSMGPIQANHTGTGLLLTLQPEKFKQVKGF 120
QY 121 GGAMTDAALNILALSPPAQNLLKSYFSEEGIGNIIRVPMASCSFIRTYTYADTPDD 180
Db 121 GGAMTDAALNILALSPPAQNLLKSYFSEEGIGNIIRVPMASCSFIRTYTYADTPDD 180
QY 181 FQHNFSLPEDTKIKIPLIHRALQAPVSLASPTWLTNGAVNGKSLKGQP 240
|||||

Db 181 FQHNFSLPEDTKIKIPLIHRALQAPVSLASPTWLTNGAVNGKSLKGQP 240
QY 241 GDIYHQTWARYFVKFLDAYAEHKLQFVAVTAENEPSAGLLSGYPFOCLGFTPEHQDFIA 300
Db 241 GDIYHQTWARYFVKFLDAYAEHKLQFVAVTAENEPSAGLLSGYPFOCLGFTPEHQDFIA 300
QY 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIADVHLYDLFLAPAK 360
Db 301 RDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIADVHLYDLFLAPAK 360
QY 361 ATLGETHRLFPNTMLFASEACVSGKFWEQSVRLGSDWDRGMQYSHSIITNLLYHVVGWTDW 420
Db 361 ATLGETHRLFPNTMLFASEACVSGKFWEQSVRLGSDWDRGMQYSHSIITNLLYHVVGWTDW 420
QY 421 NLALNPEGPNWVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPESQVRGLVASQK 480
Db 421 NLALNPEGPNWVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPESQVRGLVASQK 480
QY 481 NDLDVALMHDPDGSVAVVVLNRSSKDVPLTIKDPVAGFLETISPGYSIHITLWRRQ 536
Db 481 NDLDVALMHDPDGSVAVVVLNRSSKDVPLTIKDPVAGFLETISPGYSIHITLWRRQ 536
RESULT 4
AAE02446
ID AAE02446 standard; Protein; 536 AA.
XX
AC AAE02446;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human glucocerebrosidase (GC) #2.
XX
KW Human; adeno-associated viral expression vector; AAV; gene therapy;
KW lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
KW sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG;
KW glucocerebrosidase; GC; Gaucher's disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..39
FT /label= Signal_peptide
FT Protein 40..536
FT /note= "Human mature glucocerebrosidase (GC)"
XX WO200136603-A2.
XX 25-MAY-2001.
XX 17-NOV-2000; 2000WO-US31688.
XX 17-NOV-1999; 99US-0166097.
XX 30-JUN-2000; 2000US-0215430.
PA (AVIG-) AVIGEN INC.
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI Podsakoff G, Watson G, Couto LB, Yang B;
XX WPI; 2001-343814/36.
DR N-PSDB; AAD06389.
XX
PT Use of recombinant adeno-associated virus, comprising gene encoding a
PT protein defective or missing in lysosomal storage disease, in the
PT manufacture of a medicament for treating the lysosomal storage disease
PT
XX
PS Disclosure; Page 94-96; 97pp; English.
XX
CC The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful

CC in the manufacture of a medicament for treating lysosomal storage
CC disease e.g., mucopolysaccharidoses VII (MPS VII), MPS VII (Sly Syndrome)
CC is due to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS)
CC which aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in
CC gene therapy. The present sequence is human glucocerebrosidase (GC).
CC Mutations in GC gene leads to Gaucher's disease, a lysosomal
CC storage disease. This sequence is used in AAV constructs.

XX Sequence 536 AA;

Query Match 98.8%; Score 2832; DB 22; Length 536;
Best Local Similarity 98.9%; Pred. No. 2.3e-287;
Matches 530; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MEFSSPSRECPKPLSRVSIMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
Db 1 MEFSSPSRECPKPLNDPEGAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
QY 61 YCDSFDPPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTQPEKQKVKGF 120
Db 61 YCDSFDPPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTQPEKQKVKGF 120
QY 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
Db 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
QY 181 FQLHNFSLPEEDTKLKIPLIHRALQLAQRPVSLASPWTSPWLKTNGAVNGKSLKGQP 240
Db 181 FQLHNFSLPEEDTKLKIPLIHRALQLAQRPVSLASPWTSPWLKTNGAVNGKSLKGQP 240
QY 241 GDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIA 300
Db 241 GDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIA 300
QY 301 RDLGPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
Db 301 RDLGPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAK 360
QY 361 ATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
Db 361 ATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDW 420
QY 421 NLALNPEGGPNWVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFPEGSQVRGLVASQK 480
Db 421 NLALNPEGGPNWVRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFPEGSQVRGLVASQK 480
QY 481 NDLDVALMHPDGSVAVVVVNLNRSSKDVPLTIKDPAVGFLETISPGYSIHTYLWRRQ 536
Db 481 NDLDVALMHPDGSVAVVVVNLNRSSKDVPLTIKDPAVGFLETISPGYSIHTYLWHRQ 536

RESULT 5

AAAY01595
ID AAY01595 standard; Protein; 516 AA.

XX AC AAY01595;

XX DT 17-JUN-1999 (first entry)

XX DE Human lysosomal glucocerebrosidase protein.

XX KW Human lysosomal glucocerebrosidase; glycosylated; Gaucher's disease.

XX OS Homo sapiens.

XX PN US5879680-A.

XX PD 09-MAR-1999.

XX PF 26-MAY-1995; 95US-0452398.

XX PR 23-DEC-1987; 87US-0137796.

XX PR 05-FEB-1990; 90US-0474307.

PR 06-AUG-1992; 92US-0925333.
PR 13-JAN-1994; 94US-0186256.
PR 26-MAY-1995; 95US-0452398.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PT Eliason WK, Ginns EJ, LaMarca ME, Martin B, Maysak KA;

XX WPI; 1999-203905/17.

DR N-PSDB; AAX26682.

XX New compositions containing glucocerebrosidase - comprising
PT glycosylated recombinantly-produced human glucocerebrosidase, used
PT for treating Gaucher's disease

XX Claim 1; Fig 1A-D; 10pp; English.

XX The present sequence represents a human lysosomal glucocerebrosidase.
CC The specification describes a composition containing glycosylated
CC recombinantly produced human glucocerebrosidase which can be used
CC for treating Gaucher's disease.

XX SQ Sequence 516 AA;

Query Match 96.4%; Score 2762; DB 20; Length 516;
Best Local Similarity 100.0%; Pred. No. 4.6e-280;
Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 MAGSLTGLLLLQAVSWASGARPCIPKSGYSSVVCVNCATYCDSDPTFPALGTFSTRYE 80
Db 1 MAGSLTGLLLLQAVSWASGARPCIPKSGYSSVVCVNCATYCDSDPTFPALGTFSTRYE 60
QY 81 STRSGRRMELSMGPIQANHTGTGLLLTQPEKQKVKFGGAMTDAALNIALSPPAQ 140
Db 61 STRSGRRMELSMGPIQANHTGTGLLLTQPEKQKVKFGGAMTDAALNIALSPPAQ 120
QY 141 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDDFQLHNFSLPEEDTKLKIPLI 200
Db 121 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDDFQLHNFSLPEEDTKLKIPLI 180
QY 201 HRAIQLAQRVSVLLASPWTSPWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYA 260
Db 181 HRAIQLAQRVSVLLASPWTSPWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYA 240
QY 261 EHKLOFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPPTLANSTHNNVRLML 320
Db 241 EHKLOFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPPTLANSTHNNVRLML 300
QY 321 DDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 380
Db 301 DDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 360
QY 381 CVGSKFEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDNLANLNEGGPNWVRNFVDS 440
Db 361 CVGSKFEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDNLANLNEGGPNWVRNFVDS 420
QY 441 IIVDITKDTFYKQPMFYHLGHFSKFPEGSQVRGLVASQKNDLDAVALMHPDGSVAVVV 500
Db 421 IIVDITKDTFYKQPMFYHLGHFSKFPEGSQVRGLVASQKNDLDAVALMHPDGSVAVVV 480
QY 501 NRSSKDVPLTIKDPAVGFLETISPGYSIHTYLWRRQ 536
Db 481 NRSSKDVPLTIKDPAVGFLETISPGYSIHTYLWRRQ 516

RESULT 6

AAAY99599
ID AAY99599 standard; Protein; 516 AA.

XX AC AAY99599;

XX DT 08-SEP-2000 (first entry)

XX

DE Human lysosomal glycoprotein glucocerebrosidase (GCS).

XX Human; lysosomal glycoprotein glucocerebrosidase; Gaucher's disease;

KW GCS; beta-D-glucosyl-N-acylsphingosine glycohydrolase; neuroprotective;

KW hepatotropic; gene therapy.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein /label= Signal_peptide

FT /label= Mature_protein

XX US6074864-A.

XX 13-JUN-2000.

XX 13-JAN-1994; 94US-0186256.

XX 23-DEC-1987; 87US-0137796.

PR 05-FEB-1990; 90US-0474307.

PR 06-AUG-1992; 92US-0925333.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA Ginns EI, Maysak KA, Eliason WK, Lamarca ME, Martin B;

XX WPI; 2000-430396/37.

DR N-PSDB; AAA48444.

XX New glycosylated human glucocerebrosidase useful for treating Gaucher's

PT disease, is produced by transfecting Spodoptera frugiperda cells with

PT baculovirus containing cDNA encoding glucocerebrosidase -

XX Claim 4; Fig 1; 12pp; English.

PS The present sequence is the human lysosomal glycoprotein

XX glucocerebrosidase (GCS, beta-D-glucosyl-N-acylsphingosine

CC glycohydrolase). Mutation or deficiency of the GCS gene leads to

CC Gaucher's disease. In the present invention, a Spodoptera frugiperda

CC insect cell infected with a recombinant baculovirus vector containing a

CC the coding sequence for the present protein has been developed. This cell

CC line is useful for large-scale production of GCS, which could be used to

CC treat Gaucher's disease via gene therapy.

XX

SQ Sequence 516 AA;

Query Match 96.4%; Score 2762; DB 21; Length 516;

Best Local Similarity 100.0%; Pred. No. 4.6e-280;

Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 MAGSLTGLLLQAVSWASGARPICIPKSFSGYSSVVCVNCATYCDSPDPTFPALGTFERYE 80

DB 1 MAGSLTGLLLQAVSWASGARPICIPKSFSGYSSVVCVNCATYCDSPDPTFPALGTFERYE 60

QY 81 STRSGRRMELSMGPIQANHTGTGLLTQLQPEQKFKQVKGFGGAMTDAALNIALSPPAQ 140

DB 61 STRSGRRMELSMGPIQANHTGTGLLTQLQPEQKFKQVKGFGGAMTDAALNIALSPPAQ 120

QY 141 NLLKSYFSEEGYNIIRVPMASCDPSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLI 200

DB 121 NLLKSYFSEEGYNIIRVPMASCDPSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLI 180

QY 201 HRALQAPVSLASPTWPTSLKTNKNGAVNGKSLKQPGDIYHQTWARYFVKFLDAYA 260

DB 181 HRALQAPVSLASPTWPTSLKTNKNGAVNGKSLKQPGDIYHQTWARYFVKFLDAYA 240

QY 261 EHKLFQWAVTAENPESAGLLSGYFPQCLGFTPEHQDFIARDLGTLANSTHNNVRLML 320

DB 241 EHKLFQWAVTAENPESAGLLSGYFPQCLGFTPEHQDFIARDLGTLANSTHNNVRLML 300

QY 321 DDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 380

DE 301 DDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 360

QY 381 CVGSKFEQSVRLGSDRGDMQYSHSIITNLLYHVVGWTDNMLALNPEGGPNWVRNFVDS 440

DB 361 CVGSKFEQSVRLGSDRGDMQYSHSIITNLLYHVVGWTDNMLALNPEGGPNWVRNFVDS 420

QY 441 IIVDITKDTFYKQPMFYHLGHFSKFIPESQSRVGLVASQKNDLDAVALMHDPDGSVVVVL 500

DB 421 IIVDITKDTFYKQPMFYHLGHFSKFIPESQSRVGLVASQKNDLDAVALMHDPDGSVVVVL 480

QY 501 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLMRRQ 536

DB 481 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLMRRQ 516

RESULT 7

AAG78410

ID AAG78410 standard; Protein; 516 AA.

XX AAG78410;

AC AAG78410;

DT 13-FEB-2002 (first entry)

XX Amino acid sequence of human glucocerebrosidase.

XX Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;

KW Farber disease; Tay-Sachs disease; Niemann-Pick disease;

KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;

KW Gaucher disease; glycosylation.

XX Homo sapiens.

OS WO200177307-A2.

PN 18-OCT-2001.

XX 06-APR-2001; 2001WO-US11144.

PF 06-APR-2000; 2000US-195598P.

PR (CYTO-) CYTOCLONAL PHARM INC.

XX Berent SL;

PI WPI; 2002-041292/05.

DR N-PSDB; AAH99924.

XX Novel expression system useful for producing glucocerebrosidase for

PT treating Gaucher's disease, comprises an insect cell transformed with a

PT vector encoding glucocerebrosidase that synthesizes glucocerebrosidase

PT

XX Example 1; Page 65-67; 74pp; English.

XX This invention relates to an expression system comprising an insect

CC cell transformed with a vector encoding glucocerebrosidase (GC) that

CC synthesizes clinically effective GC. Recombinant GC is useful for

CC treating individuals with deficiencies in GC, by introducing the

CC clinically effective form of recombinant GC from insect cells into

CC individuals. The expression system is stably transformed and

CC provides a consistently higher level of expression of GC than in

CC baculovirus or mammalian cell expression systems, and proper

CC glycosylation modifications for GC, requiring no enzymatic

CC carbohydrate remodeling to be clinically effective. The expression

CC system is more effective, economical and simple for the manufacture of

CC recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack

CC of lysosomal enzymes contributes to some of the many genetically

CC inherited lysosomal storage diseases, such as Fabry disease, Farber

CC disease, Tay-Sachs disease, Niemann-Pick disease, Hunter syndrome,

CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease.

CC This amino acid sequence represents that of human glucocerebrosidase.

CC See also AAG78411 for the amino acid sequence of human

RESULT 9
AAG78411
ID AAG78411 standard; Protein; 516 AA.
XX
AC AAG78411;
XX
DT 13-FEB-2002 (first entry)
XX
DE Amino acid sequence of human glucocerebrosidase.
XX
KW Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
KW Farber disease; Tay-sachs disease; Niemann-pick disease;
KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
KW Gaucher disease; glycosylation.
XX
OS Homo sapiens.
XX
PN WO200177307-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11144.
XX
PR 06-APR-2000; 2000US-195598P.
XX
PA (CYTO-) CYTOCLONAL PHARM INC.
XX
PI Berent SL;
XX
DR WPI; 2002-041292/05.
DR N-PSDB; AAH99925.
XX
PT Novel expression system useful for producing glucocerebrosidase for
PT treating Gaucher's disease, comprises an insect cell transformed with a
PT vector encoding glucocerebrosidase that synthesizes glucocerebrosidase
PT
XX
PS Example 1; Page 70-72; 74pp; English.
XX
CC This invention relates to an expression system comprising an insect
CC cell transformed with a vector encoding glucocerebrosidase (GC) that
CC synthesizes clinically effective GC. Recombinant GC is useful for
CC treating individuals with deficiencies in GC, by introducing the
CC clinically effective form of recombinant GC from insect cells into
CC individuals. The expression system is stably transformed and
CC provides a consistently higher level of expression of GC than in
CC baculovirus or mammalian cell expression systems, and proper
CC glycosylation modifications for GC, requiring no enzymatic
CC carbohydrate remodeling to be clinically effective. The expression
CC system is more effective, economical and simple for the manufacture of
CC recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack
CC of lysosomal enzymes contributes to some of the many genetically
CC inherited lysosomal storage diseases, such as Fabry disease, Farber
CC disease, Tay-Sachs disease, Niemann-pick disease, Hunter syndrome,
CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease.
CC This amino acid sequence represents that of human
CC glucocerebrosidase. See also AAG78410 for the amino acid sequence of
CC human glucocerebrosidase, noting the difference at residue 514. This
CC sequence has histidine at this position, and AAG78410 has arginine.
XX
SQ Sequence 516 AA;
Query Match 96.2%; Score 2757; DB 23; Length 516;
Best Local Similarity 99.8%; Pred. NO. 1.6e-279;
Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 21 MAGSLTGLLLQAVSWAGARPCIPKSFSGYSSVVCNATYCDSPDPTFPALGTFSTRYE 80
DB 1 MAGSLTGLLLQAVSWAGARPCIPKSFSGYSSVVCNATYCDSPDPTFPALGTFSTRYE 60
QY 81 STRSGRRMELSMGPIQANHTGTGLLLTLOPEQKQKVKGFGGAMTDAALNILALSPPAQ 140
|||||

Db 61 STRSGRRMELSMGPIQANHTGTGLLLTLOPEQKQKVKGFGGAMTDAALNILALSPPAQ 120
QY 141 NLLLSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLI 200
|||||
Db 121 NLLLSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLI 180
|||||
QY 201 HRALQLAQRPVSLASPTWLTNGAVNGKSLKQPGDIYHQTWARYFVKFLDAYA 260
|||||
Db 181 HRALQLAQRPVSLASPTWLTNGAVNGKSLKQPGDIYHQTWARYFVKFLDAYA 240
|||||
QY 261 EHKLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPTLANSTHHNVRLLML 320
|||||
Db 241 EHKLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPTLANSTHHNVRLLML 300
|||||
QY 321 DDQRLLLPHWAKVVLTDPEAAKYVHGIYVHGYLDLFLAPAKATLGETHRLFPNTMLFASEA 380
|||||
Db 301 DDQRLLLPHWAKVVLTDPEAAKYVHGIYVHGYLDLFLAPAKATLGETHRLFPNTMLFASEA 360
|||||
QY 381 CVGSKFWEQSVRLGSDRGMQYSHSIIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDS 440
|||||
Db 361 CVGSKFWEQSVRLGSDRGMQYSHSIIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDS 420
|||||
QY 441 IIVDITKDTFYKQPMFYHLGHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSVAVVVL 500
|||||
Db 421 IIVDITKDTFYKQPMFYHLGHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSVAVVVL 480
|||||
QY 501 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLRWRQ 536
Db 481 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLRWRQ 516
RESULT 10
AAP81987
ID AAP81987 standard; protein; 516 AA.
XX
AC AAP81987;
XX
DT 17-DEC-2001 (updated)
DT 12-OCT-1990 (first entry)
XX
DE Recombinant human lysosomal glucocerebrosidase.
XX
KW lysosomal glycoprotein glucocerebrosidase ; Gaucher's disease; ss.
XX
OS synthetic.
FH Key Location/Qualifiers
FT Misc-difference 489..489 /label=old_seq(pro)
FT Misc-difference 514..514 /label=old_seq(His)
XX
PN USN7137796-N.
XX
PD 02-AUG-1988.
XX
PF 23-DEC-1987; 87US-0192026.
XX
PR 23-DEC-1987; 87US-0137796.
XX
PA (USSH) US DEPT HEALTH AND HUMAN SERVICES.
XX
PI Ginns EI;
XX
DR WPI; 1988-264274/37.
DR N-PSDB; AAN80113.
XX
PT Recombinant glucocerebrosidase -
PT useful for treating Gaucher's disease
XX
PS Disclosure; ; 2pp; English.
XX
CC The protein produced by the baculovirus expression system lacks

QY 289 GFTPEHQDFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIA 348
|||||
Db 345 GFTPEHQDFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGIA 404
|||||
QY 349 VHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRGMQYSHSIIT 408
|||||
Db 405 VHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRGMQYSHSIIT 464
|||||
QY 409 NLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIVDTIKDTFYKQPMFYHLGHFSKFIPE 468
|||||
Db 465 NLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIVDTIKDTFYKQPMFYHLGHFSKFIPE 524
|||||
QY 469 GSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAVGFLETISPGYSI 528
|||||
Db 525 GSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAVGFLETISPGYSI 584
|||||
QY 529 HTYLWRRQ 536
|||||
Db 585 HTYLWRRQ 592
|||||

RESULT 12
AAW07885
ID AAW07885 standard; protein; 498 AA.
XX
AC AAW07885;
XX
DT 28-JAN-1997 (first entry)
XX
DE Remodelled, recombinant glucocerebrosidase (r-GCR).
XX
KW Remodelling; glycosylation; glucocerebrosidase; GCR; anaemia;
KW lysozyme deficiency; glycolipid hydrolysis; Gaucher's disease;
KW splenomegaly; hepatomegaly; skeletal disorder; thrombocytopenia;
KW placenta-derived; targetting.
XX
OS Synthetic.
XX
FH key Location/Qualifiers
FT Misc-difference 495
FT /label= substitution
FT /note= "His replaces Arg, found at this posn. in
FT naturally occurring placental GCR (p-GCR)"
XX
PN US5549892-A.
XX
PD 27-AUG-1996.
XX
PF 23-DEC-1988; 88US-0289589.
XX
PR 21-JUN-1993; 93US-0080855.
PR 23-DEC-1988; 88US-0289589.
PR 22-DEC-1989; 89US-0455507.
PR 21-AUG-1991; 91US-0748283.
XX
PA (GENZ) GENZYME CORP.
XX
PI Friedman B, Hayes M;
XX
DR WPI; 1996-401555/40.
XX
PT Re-modelled glucocerebrosidase has increased fucose and N-acetyl
PT glucosamine residues - effectively alleviates clinical symptoms of
PT Gaucher's disease at a lower dose than naturally occurring enzyme
XX
PS Example 1; Column 9-12; 9pp; English.
XX
CC AAW07885 is a recombinant version of the commercially available
CC remodelled glucocerebrosidase, p-GCR (derived from placental tissue).
CC The recombinant glucocerebrosidase (r-GCR) has a His-Arg substitution
CC at posn. 495 and is remodelled to increase the level of glycosylation
CC of the enzyme, having four complex carbohydrate chains and increased

CC fucose and N-acetyl glucosamine (Glc-Nac) residues compared to p-GCR.
CC The remodelled r-GCR has improved pharmacokinetics, including
CC approximately a two-fold increase in affinity for its target cells
CC (Kupffer cells). Remodelled r-GCR is useful in treating patients with
CC Gaucher's disease.

XX Sequence 498 AA;

Query Match 93.0%; Score 2665; DB 17; Length 498;
Best Local Similarity 99.8%; Pred. No. 6.4e-270;
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 40 ARPCIPKSFYSSVVCNATYCDSPDPTFPALGTFSTRSGRMELSMGPIQANH 99
|||||
Db 1 ARPCIPKSFYSSVVCNATYCDSPDPTFPALGTFSTRSGRMELSMGPIQANH 60
|||||
QY 100 TGTGLLLTLQPEQKQKVGKGGGAMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIR 159
|||||
Db 61 TGTGLLLTLQPEQKQKVGKGGGAMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIR 120
|||||
QY 160 VPMASCDFSIRTYADTPDDFQLHNFSLPEEDTKIKPLIHRALQLAQRPVSLASPWT 219
|||||
Db 121 VPMASCDFSIRTYADTPDDFQLHNFSLPEEDTKIKPLIHRALQLAQRPVSLASPWT 180
|||||
QY 220 SPTWLKTNGAVNGKSLKQPGDIHQWTWARYFVKELDAYAEHKLQFWAVTAENEPSAGL 279
|||||
Db 181 SPTWLKTNGAVNGKSLKQPGDIHQWTWARYFVKELDAYAEHKLQFWAVTAENEPSAGL 240
|||||
QY 280 LSGYPFQCLGFTPEHQDFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 339
|||||
Db 241 LSGYPFQCLGFTPEHQDFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 300
|||||
QY 340 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 399
|||||
Db 301 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 360
|||||
QY 400 MQYSHSIITNLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIVDTIKDTFYKQPMFYHL 459
|||||
Db 361 MQYSHSIITNLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIVDTIKDTFYKQPMFYHL 420
|||||
QY 460 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAVGF 519
|||||
Db 421 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPLTIKDPVAVGF 480
|||||
QY 520 ETISPGYSIHTYLWRRQ 536
|||||
Db 481 ETISPGYSIHTYLWRRQ 497
|||||

RESULT 13

AAU05696
ID AAU05696 standard; Protein; 497 AA.
XX

AC AAU05696;

DT 24-OCT-2001 (first entry)

XX Human Mature glucocerebrosidase, GCB.

DE Human; glucocerebrosidase; GCB; lysosomal storage disease;
KW Gaucher's disease; Fabry's disease; Farber's disease;
KW G_m1 gangliosidosis; Tay-Sach's disease; Niemann-Pick disease;
KW Shindler disease; Hunter syndrome; Sly syndrome; Huler syndrome;
KW Scheie syndrome.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 19 /note= "N-glycosylated"

FT Modified-site 59 /note= "N-glycosylated"

FT Modified-site 146

FT /note= "N-glycosylated"
FT Misc-difference 259
FT /note= "Encoded by TTA"
FT Modified-site 270
FT /note= "N-glycosylated"
FT Misc-difference 495
FT /label= OTHER
FT /note= "OTHER= Arg or His"
XX
PN
XX
PD
XX
PF
XX
PR
PR
PR
PR
PR
XX
PA
XX
PI
XX
XX
DR
DR
XX
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

W0200149830-A2.
12-JUL-2001.
29-DEC-2000; 2000WO-DK00743.
30-DEC-1999; 99DK-0001891.
02-JUN-2000; 2000DK-0000865.
02-JUN-2000; 2000DK-0000866.
30-JUN-2000; 2000DK-0001027.
(MAXY-) MAXYGEN APS.
Okkels JS, Jensen AD, Halkier T, Jensen RB, Schambye HT;
WPI; 2001-465259/50.
N-PSDB; AAS10863.
Improved lysosomal enzymes and lysosomal enzyme activators useful for
treating Gaucher's disease -
Claim 20; Page 96; 97pp; English.
The sequence is the mature human lysosomal enzyme glucocerebrosidase,
GCB. GCB is the enzyme involved in Gaucher's disease, a lysosomal
storage disease. The invention relates to introducing new glycosylation
sites into lysosomal enzymes/activators like GCB to improve their
bioactivity. The novel polypeptides are used for the prevention and
treatment of Gaucher's disease, Fabry's disease, Farber's disease,
G_m_1 gangliosidosis, Tay-Sach's disease, Niemann-Pick disease, Shindler
disease, Hunter syndrome, Sly syndrome, Hurler and Huler/Scheie
syndromes.

Query Match 93.0%; Score 2664; DB 22; Length 497;
Best Local Similarity 99.8%; Pred. No. 8.2e-270;
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 40 ARPCIPKSGYSSVVCNATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 99
Db 1 ARPCIPKSGYSSVVCNATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 60
QY 100 TGTGLLLTLQPEQKFKQKVGFGGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 159
Db 61 TGTGLLLTLQPEQKFKQKVGFGGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 120
QY 160 VPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRPVSLASPWT 219
Db 121 VPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRPVSLASPWT 180
QY 220 SPTWLKTNAGVNGKSLKGQPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 279
Db 181 SPTWLKTNAGVNGKSLKGQPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 240
QY 280 LSGYPFCQLGFTPEHQRFIARDLPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 241 LSGYPFCQLGFTPEHQRFIARDLPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 300
QY 340 AAKYVHGIAVHWYLDLAPAKATIGETHRLFPNTMLFASACVGSKFWEQSVRLGSWDRG 399
Db 301 AAKYVHGIAVHWYLDLAPAKATIGETHRLFPNTMLFASACVGSKFWEQSVRLGSWDRG 360
QY 400 MQYSHSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDTKDTFYKQPMFYHL 459

Db 361 MQYSHSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDTKDTFYKQPMFYHL 420
QY 460 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPPLTIKDPAVGFL 519
Db 421 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVVVVLNRSSKDVPPLTIKDPAVGFL 480
QY 520 ETISPGYSIHTYLRWRRQ 536
Db 481 ETISPGYSIHTYLRWRRQ 497

RESULT 14
AAU86020
ID AAU86020 standard; Protein; 497 AA.
XX
AC AAU86020;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human glucocerebrosidase.
XX

KW Human; glucocerebrosidase; GCB; antidiabetic; cytostatic; cancer;
KW antiinflammatory; dermatological; immunosuppressive; antirheumatic;
KW antiarthritic; antithyroid; detergent; textile; food product; vaccine;
KW agrochemical; feed product; viral pathogen; parasite; autoantigen;
KW autoimmune disease; insulin-dependent diabetes mellitus; gene therapy;
KW autoimmune thyroid disorder; inflammatory condition; sperm antigen.
XX Homo sapiens.
XX
PN W0200202597-A2.
XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-DK00459.
PR 30-JUN-2000; 2000DK-0001027.
PR 14-JUL-2000; 2000DK-0001092.
PR 29-DEC-2000; 2000WO-DK00743.
PR 09-FEB-2001; 2001WO-DK00090.
XX
PA (MAXY-) MAXYGEN APS.
PA (MAXY-) MAXYGEN HOLDINGS LTD.

XX Okkels JS, Jensen AD, Van Den Hazel B;
XX
DR WPI; 2002-195745/25.
DR N-PSDE; ABK40235.

PT Novel glycosylated polypeptides having improved properties as compared
PT to polypeptide of interest e.g. improved shelf-life, stability, and
XX reduced renal clearance, comprise an additional glycosylation site -
PS
XX Example 2; Page 90-92; 136pp; English.

CC The invention relates to novel glycosylated polypeptides (I) into which
CC an additional glycosylation site (GS) has been introduced. (I) comprises
CC the primary structure: NH2-X-Pp-COOH or NH2-Px-X-Py-COOH where:
CC X = a peptide addition comprising or contributing to a GS; Pp = a
CC polypeptide; Px = an N-terminal part of Pp; and Py = a C-terminal of Pp.
CC Also described is a method of improving one or more selected properties
CC of Pp, by preparing a nucleotide sequence encoding a polypeptide with the
CC primary structure NH2-X-Pp-COOH, where X is a peptide addition comprising
CC or contributing to a GS that is capable of conferring the selected
CC improved properties to Pp, expressing the nucleotide sequence in a
CC suitable host cell, optionally conjugating the polypeptide to a
CC second non-peptide moiety and recovering the polypeptide. The
CC glycosylated polypeptides have improved properties compared to the
CC polypeptide of interest. Pp is chosen from an antibody or its fragment,
CC plasma protein, erythrocyte or thrombocyte protein, cytokine, growth
CC factor, profibrinolytic protein, protease inhibitor, antigen, enzyme,
CC ligand, receptor, or a hormone, or is a microbial enzyme chosen from

CC protease, amylase, amyloglucosidase, pectinase, lipase and cutinase. Pp
CC can be a therapeutic polypeptide useful in human or veterinary therapy, a
CC diagnostic polypeptide used in diagnostics, an industrial polypeptide
CC useful in the manufacture of products such as detergents, household
CC articles, textile, food products, agrochemicals, feed products,
CC polypeptides or proteins associated with viral pathogens, antigens from
CC parasites, autoantigens associated with autoimmune diseases such as
CC insulin-dependent diabetes mellitus; autoimmune thyroid disorders;
CC inflammatory conditions; sperm antigens; and bacterial and cancer
CC antigens. When the polypeptide is an in vivo glycosylated
CC polypeptide which does not comprise any other type of non-peptide
CC moiety, a nucleotide sequence encoding the polypeptide can be used in
CC gene therapy. When the polypeptide Pp is an antigen, the polypeptide
CC can be provided in the form of a vaccine. AAU86020-AAU86128 represent
CC the peptide addition sequences, and related protein sequences of
CC the invention.
XX
SQ Sequence 497 AA;

Query Match 93.0%; Score 2664; DB 23; Length 497;
Best Local Similarity 99.8%; Pred. No. 8.2e-270;
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 40 ARPCIPKSFYSSVVCVNCATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 99
Db 1 ARPCIPKSFYSSVVCVNCATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 60
QY 100 TGTGLLLTLOPEQKFKQVKVGGAMTDAALNIALSPPAQNLLKSYFSEEGYNIIR 159
Db 61 TGTGLLLTLOPEQKFKQVKVGGAMTDAALNIALSPPAQNLLKSYFSEEGYNIIR 120
QY 160 VPMASCDFSIRTYTYADTPDDFQLHNSLPEDTKLKIPLIHRALQLAQRPSVLLASPWT 219
Db 121 VPMASCDFSIRTYTYADTPDDFQLHNSLPEDTKLKIPLIHRALQLAQRPSVLLASPWT 180
QY 220 SPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 279
Db 181 SPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 240
QY 280 LSGYPFQCLGTFPEHQDFIARDLPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 241 LSGYPFQCLGTFPEHQDFIARDLPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 300
QY 340 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 399
Db 301 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 360
QY 400 MQYSHSIITNLYHVVGWTDNNLALNPEGPNWVNFVDSPIIVDITKDTFYKOPMFYHL 459
Db 361 MQYSHSIITNLYHVVGWTDNNLALNPEGPNWVNFVDSPIIVDITKDTFYKOPMFYHL 420
QY 460 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVAVVVLNRSKDVPLTIKDPVAGFL 519
Db 421 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVAVVVLNRSKDVPLTIKDPVAGFL 480
QY 520 ETISPGYSIHTYLRWRRQ 536
Db 481 ETISPGYSIHTYLRWRRQ 497

RESULT 15
AAU05668
ID AAU05668 standard; Protein; 497 AA.
XX
AC AAU05668;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human glucocerebrosidase, GCB, mutant K194N.
XX
KW Human; glucocerebrosidase; GCB; lysosomal storage disease;
KW Gaucher's disease; Fabry's disease; Farber's disease;
KW G_m_1 gangliosidosis; Tay-Sach's disease; Niemann-Pick disease;

KW Shindler disease; Hunter syndrome; Sly syndrome; Huler syndrome;
KW Schele syndrome; mutant; mutein; K194N.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 19 /note= "N-glycosylated"
FT Modified-site 59 /note= "N-glycosylated"
FT Modified-site 146 /note= "N-glycosylated"
FT Modified-site 194 /note= "N-glycosylated"
FT Misc-difference 194 /note= "Wild-type Lys changed to Asn"
FT Modified-site 270 /note= "N-glycosylated"
FT Misc-difference 495 /label= OTHER
FT /note= "OTHER= Arg or His"
XX
PN WO200149830-A2.
XX 12-JUL-2001.
XX 29-DEC-2000; 2000WO-DK00743.
XX 30-DEC-1999; 99DK-0001891.
PR 02-JUN-2000; 2000DK-0000865.
PR 02-JUN-2000; 2000DK-0000866.
PR 30-JUN-2000; 2000DK-0001027.
XX (MAXY-) MAXYGEN APS.
XX Okkels JS, Jensen AD, Halkier T, Jensen RB, Schambye HT;
PI WPI; 2001-465259/50.
DR Improved lysosomal enzymes and lysosomal enzyme activators useful for
XX treating Gaucher's disease -
PS Claim 20; Page -; 97pp; English.
XX The sequence is a mutant of human lysosomal enzyme
CC glucocerebrosidase, GCB, which has an introduced glycosylation site.
CC GCB is the enzyme involved in Gaucher's disease, a lysosomal
CC storage disease. The invention relates to introducing new glycosylation
CC sites into lysosomal enzymes/activators like GCB to improve their
CC bioactivity. The novel polypeptides are used for the prevention and
CC treatment of Gaucher's disease, Fabry's disease, Farber's disease,
CC G_m_1 gangliosidosis, Tay-Sach's disease, Niemann-Pick disease, Shindler
CC disease, Hunter syndrome, Sly syndrome, Huler and Huler/Schele
CC syndromes.
CC Note: The present sequence is not displayed in the specification
CC but was created from the GCB sequence shown on page 96.
XX
SQ Sequence 497 AA;

Query Match 92.8%; Score 2659; DB 22; Length 497;
Best Local Similarity 99.6%; Pred. No. 2.7e-269;
Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 40 ARPCIPKSFYSSVVCVNCATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 99
Db 1 ARPCIPKSFYSSVVCVNCATYCDSDPPTFPALGTFSTRSRRMELSMGPIQANH 60
QY 100 TGTGLLLTLOPEQKFKQVKVGGAMTDAALNIALSPPAQNLLKSYFSEEGYNIIR 159
Db 61 TGTGLLLTLOPEQKFKQVKVGGAMTDAALNIALSPPAQNLLKSYFSEEGYNIIR 120
QY 160 VPMASCDFSIRTYTYADTPDDFQLHNSLPEDTKLKIPLIHRALQLAQRPSVLLASPWT 219

Db 121 VPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHRAQLAQRVPSLLASPWT 180
QY 220 SPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 279
Db 181 SPTWLKTNGAVNGSLKGQPGDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 240
QY 280 LSGYPFQCLGFTPEHQDFIARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 241 LSGYPFQCLGFTPEHQDFIARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKVVLTDPE 300
QY 340 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 399
Db 301 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRG 360
QY 400 MOYSHSIITNLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIIVDITKDTFYKQPMFYHL 459
Db 361 MOYSHSIITNLLYHVVGWTDWNLALNPEGGPNVVRNFVDSPIIIVDITKDTFYKQPMFYHL 420
QY 460 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVAVVVVNLNRSSKDVPLTIKDPAVGFL 519
Db 421 GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVAVVVVNLNRSSKDVPLTIKDPAVGFL 480
QY 520 ETISPGYSIHTYLWRRQ 536
Db 481 ETISPGYSIHTYLWXRQ 497

Search completed: February 20, 2003, 11:18:10
Job time : 41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:19:22 ; Search time 14 Seconds
(without alignments)
978.157 Million cell updates/sec

Title: US-10-024-197-25
Perfect score: 2865
Sequence: 1 MEFSSPSRECPKPLSRVSI.....GFLETISPGYSIHTYLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep: *
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep: *
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep: *
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep: *
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep: *
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep: *
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep: *
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep: *
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep: *
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep: *
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep: *
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep: *
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep: *
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 2675 | 93.4 | 592 | 10 | US-09-753-126-4 |
| 2 | 2664 | 93.0 | 497 | 10 | US-09-753-126-1 |
| 3 | 92 | 3.2 | 399 | 10 | US-09-815-242-13600 |
| 4 | 91.5 | 3.2 | 427 | 9 | US-10-062-523-16 |
| 5 | 89 | 3.1 | 740 | 9 | US-10-051-909-37 |
| 6 | 88.5 | 3.1 | 4303 | 9 | US-09-904-968A-2 |
| 7 | 88 | 3.1 | 556 | 10 | US-09-771-161A-245 |
| 8 | 87.5 | 3.1 | 551 | 10 | US-09-897-214-8 |
| 9 | 86.5 | 3.0 | 600 | 10 | US-09-801-368-158 |
| 10 | 85.5 | 3.0 | 731 | 9 | US-09-738-626-4854 |
| 11 | 85 | 3.0 | 615 | 10 | US-09-815-242-13747 |
| 12 | 84.5 | 2.9 | 667 | 9 | US-09-759-056-2 |
| 13 | 84.5 | 2.9 | 667 | 9 | US-09-901-812-2 |
| 14 | 84.5 | 2.9 | 667 | 9 | US-10-227-884-80 |
| 15 | 84.5 | 2.9 | 728 | 9 | US-10-106-092-4 |
| 16 | 84 | 2.9 | 1233 | 9 | US-09-738-626-4312 |
| 17 | 83 | 2.9 | 341 | 9 | US-09-738-626-6673 |
| 18 | 83 | 2.9 | 791 | 9 | US-09-908-193-18 |
| 19 | 82.5 | 2.9 | 658 | 9 | US-09-759-056-5 |
| | | | | | Sequence 4, Appli |
| | | | | | Sequence 1, Appli |
| | | | | | Sequence 13600, A |
| | | | | | Sequence 16, Appl |
| | | | | | Sequence 37, Appl |
| | | | | | Sequence 2, Appli |
| | | | | | Sequence 245, App |
| | | | | | Sequence 8, Appli |
| | | | | | Sequence 158, App |
| | | | | | Sequence 4854, Ap |
| | | | | | Sequence 13747, A |
| | | | | | Sequence 2, Appli |
| | | | | | Sequence 2, Appli |
| | | | | | Sequence 80, Appl |
| | | | | | Sequence 4, Appli |
| | | | | | Sequence 4312, Ap |
| | | | | | Sequence 6673, Ap |
| | | | | | Sequence 18, Appl |
| | | | | | Sequence 5, Appli |

| | | | | | | |
|----|------|-----|------|----|--------------------|-------------------|
| 20 | 82.5 | 2.9 | 658 | 9 | US-09-901-812-5 | Sequence 5, Appli |
| 21 | 81.5 | 2.8 | 953 | 9 | US-10-118-984-43 | Sequence 43, Appl |
| 22 | 81.5 | 2.8 | 953 | 10 | US-09-728-721-43 | Sequence 81, Appl |
| 23 | 81.5 | 2.8 | 2802 | 9 | US-09-808-602-81 | Sequence 6846, Ap |
| 24 | 80.5 | 2.8 | 375 | 9 | US-09-738-626-6846 | Sequence 2, Appli |
| 25 | 80.5 | 2.8 | 803 | 9 | US-09-801-220-2 | Sequence 5038, Ap |
| 26 | 80.5 | 2.8 | 828 | 9 | US-09-738-626-5038 | Sequence 518, App |
| 27 | 80.5 | 2.8 | 1137 | 9 | US-10-174-590-518 | Sequence 518, App |
| 28 | 80.5 | 2.8 | 1137 | 9 | US-10-176-758-518 | Sequence 518, App |
| 29 | 80.5 | 2.8 | 1137 | 9 | US-10-175-737-518 | Sequence 518, App |
| 30 | 80.5 | 2.8 | 1137 | 9 | US-10-173-706-518 | Sequence 518, App |
| 31 | 80.5 | 2.8 | 1137 | 9 | US-10-175-738-518 | Sequence 518, App |
| 32 | 80.5 | 2.8 | 1137 | 9 | US-10-175-752-518 | Sequence 518, App |
| 33 | 80.5 | 2.8 | 1137 | 9 | US-10-176-482-518 | Sequence 518, App |
| 34 | 80.5 | 2.8 | 1137 | 9 | US-10-176-757-518 | Sequence 518, App |
| 35 | 80.5 | 2.8 | 1137 | 9 | US-10-176-913-518 | Sequence 518, App |
| 36 | 80.5 | 2.8 | 1137 | 9 | US-10-180-552-518 | Sequence 518, App |
| 37 | 80.5 | 2.8 | 1137 | 9 | US-10-180-557-518 | Sequence 518, App |
| 38 | 80.5 | 2.8 | 1137 | 9 | US-10-173-700-518 | Sequence 518, App |
| 39 | 80.5 | 2.8 | 1137 | 9 | US-10-174-572-518 | Sequence 518, App |
| 40 | 80.5 | 2.8 | 1137 | 9 | US-10-174-579-518 | Sequence 518, App |
| 41 | 80.5 | 2.8 | 1137 | 9 | US-10-174-582-518 | Sequence 518, App |
| 42 | 80.5 | 2.8 | 1137 | 9 | US-10-174-588-518 | Sequence 518, App |
| 43 | 80.5 | 2.8 | 1137 | 9 | US-10-175-739-518 | Sequence 518, App |
| 44 | 80.5 | 2.8 | 1137 | 9 | US-10-175-740-518 | Sequence 518, App |
| 45 | 80.5 | 2.8 | 1137 | 9 | US-10-175-743-518 | Sequence 518, App |

ALIGNMENTS

RESULT 1
US-09-753-126-4
; Sequence 4, Application US/09753126
; Patent No. US20020127219A1
; GENERAL INFORMATION:
; APPLICANT: OKKELS, JENS SIGURD
; APPLICANT: JENSEN, ANNE DAM
; APPLICANT: HALKIER, TORBEN
; APPLICANT: JENSEN, RIKKE BOLDING
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME
; TITLE OF INVENTION: ACTIVATORS
; FILE REFERENCE: 31-0006000US
; CURRENT APPLICATION NUMBER: US/09/753,126
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PA 1999 01891
; PRIOR FILING DATE: 1999-12-30
; PRIOR APPLICATION NUMBER: 60/174,652
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PA 200 00865
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/210,984
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 60/211,124
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: PA 2000 01027
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/217,497
; PRIOR FILING DATE: 2000-07-11
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: SapC-linker-GCB polypeptide
US-09-753-126-4

Query Match 93.4%; Score 2675; DB 10; Length 592;
Best Local Similarity 92.3%; Pred. No. 6.4e-244;
Matches 506; Conservative 8; Mismatches 14; Indels 20; Gaps 2;


```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13600
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13600

Query Match          3.2%; Score 92; DB 10; Length 399;
Best Local Similarity 20.4%; Pred. No. 1.3;
Matches 73; Conservative 53; Mismatches 135; Indels 96; Gaps 15;

QY 108 LQPEQKQKV-----KGF-----GGAMTDA--LNILALSPPAQNLLKSY 147
Db 4  MQMPSEFQKALPVLKIKKEAGFEAYFVGSGVRDALLHSPHVDIATSSYPEETKQIFPR 63

QY 148 FSEEGIGYNIIRVPMASCDFSIRTYADTDDFQLHNF-----SLPEEDTKLKIPLIHR 202
Db 64 TADIGIEHGTVLVDGDEEYEVTTERTEDVYDYRRPSAVSFVRSLEEDLKRR-----116

QY 203 ALQLAQRPVSLASPTWLTNGCAVNGKSLKQPGDIYHQWARYFVKFLDAYAEH 262
Db 117 -----DFTVNAFALDET-----GEIVDLFHG-----LEDLEKQ 144

QY 263 KLFQWAVTAE--NEPSAGLLSGYPFQC-LGFTPEHQORDFIARDLGPTLANSTHNNVRLLM 319
Db 145 VLRAVGVASERFNDALRIMRGFRFQASLGFALEPETFKAMKTLTPLLKISVE--RTFV 202

QY 320 LDDQRLLLPHWAKVVLTDPEAAKYVHGIAVHWYLDPLAPAKATLGETHRLFPNTMLF--- 376
Db 203 EFDRKLLAPFWRRLGLASMIQSQAY-----DYLPMASSQDKLNRLFDLEDTFFE 252

QY 377 ASEACVGSKFWEQSVR-----LGSDRGMQYSHSIITNLLYHVVGWTDNLANLNEP 428
Db 253 SSEQAWAALLWALEIENASQSLKSWKTSRQFAKQVQDILLI-----LALRENG 300

RESULT 4
US-10-062-523-16
; Sequence 16, Application US/10062523
; Publication No. US20030022327A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Dendritic Enriched Secreted Lymphocyte Activation
; FILE REFERENCE: PF448P2
; CURRENT APPLICATION NUMBER: US/10/062,523
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/267,523
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: PCT/US00/21130
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/190,062
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 09/369,248
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: PCT/US99/02415
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/244,110
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 60/078,572
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/073,962
; PRIOR FILING DATE: 1998-02-06

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1-211)
; OTHER INFORMATION: Recombinant extracellular domain of D-SLAM from Q24-D233
; NAME/KEY: DOMAIN
; LOCATION: (212-427)
; OTHER INFORMATION: C-terminal Histidine tag
US-10-062-523-16

Query Match          3.2%; Score 91.5; DB 9; Length 427;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 95; Conservative 46; Mismatches 134; Indels 195; Gaps 27;

QY 68 PTFPALGTFESR-----YESTRSGR-----RMELSMGPIQANHTGTGLLLTLQPEKQK 116
Db 33 PSEELLATFFRGSLLETLYHSRFLGRAQLHSNLSLELGPESGDSG-----77

QY 117 VKFGGAMTDAALNILALSPPAQNLLKSYFSEEGIGYNIIRV-----PMASCD 167
Db 78 --NFSVLMVDTRG-----QPWTQTLLQKVY---DAVPRPVVQVFIIVERDAQPSKTCQV 126

QY 168 -----SIRTYTYA---DTPDDDFQLHNFSLPEEDTKLKIPLIHRALQLA-----QRPVS 212
Db 127 FLSCWAPNISEITYSWRRETTMDFGMEPHSLFTDQVLSISLPGDRDVAIVSIVSNPVS 186

QY 213 ---LLASPTWLTNGCAVNGKSLKQ-----PG-DIYHQTWARYF 252
Db 187 WDLATVTPWDS---CHHEAAPGKASYKDQVLSKVGSSVLLVAARPPGFQVREAIWR---239

QY 253 VKFLDAYAEHKLQFWAVTAENEPASAGLSG-----YFQCLGFTPEHQDFIARDL 303
Db 240 -----SLW-----PSEELLATFFRGSLLETLYHSRFLGRAQLHSN--LSLEL 278

QY 304 GPTLANSTHNNVRLMLDDQ-----RLL 326
Db 279 GP-LESQDSGNFSVLMVDTRGQPWTQTLLQKVYDAVPRPVVQVFIIVERDAQPSKTCQV 337

QY 327 LPHWAKVVLTDPEAAKYVHGIAVHWYLDPLAPAKATLG-ETHRLFPNTMLFASEACVGSK 385
Db 338 LSCWA-----PNISEITYSWRRETTMDF-----GMEPHSLFTDQVVL-----374

RESULT 5
US-10-051-909-37
; Sequence 37, Application US/10051909
; Publication No. US20020199217A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Helentjaris, Tim
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Tingey, Scott
; TITLE OF INVENTION: Plant Sugar Transport Proteins
; FILE REFERENCE: B01163 US CIP
; CURRENT APPLICATION NUMBER: US/10/051,909
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/083,044
; PRIOR FILING DATE: April 24, 1998
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 740
; TYPE: PRT
```


| | | | | |
|-----------------------|------------------|----------------|------------|-------------|
| Query Match | 3.0%; | Score 85; | DB 10; | Length 615; |
| Best Local Similarity | 27.5%; | Pred. No. 11; | | |
| Matches 39; | Conservative 16; | Mismatches 67; | Indels 20; | Gaps 6; |

```

; GENERAL INFORMATION.
;
; APPLICANT: Pennlica, Diane
; APPLICANT: Smith, Victoria
; APPLICANT: Wood, William I.
;
; TITLE OF INVENTION: NO. US20020156252A1e1 STRA6 Polypeptides
;
; FILE REFERENCE: GENENT.2827A2
;
; CURRENT APPLICATION NUMBER: US/09/759,056
; CURRENT FILING DATE: 2001-01-11
;
; PRIOR APPLICATION NUMBER: 60/197089
;
; PRIOR FILING DATE: 2000-04-14
;

```

| | | | | |
|-----------------------|------------------|-----------------|-------------|-------------|
| Query Match | 2.9%; | Score 84.5; | DB 9; | Length 667; |
| Best Local Similarity | 19.0%; | Pred. No. 14; | | |
| Matches 117; | Conservative 65; | Mismatches 202; | Indels 233; | Gaps 31; |

| | | | | | | | |
|----|-----|--|---|---------------------------|-------------------------|-------------------|-----|
| QY | 4 | SSPSRE--- | ECPK----- | ----- | PLSRVSIMAGSLTG | LLLLQAVSWAS-GAR-- | 41 |
| | | : : : : | : : | : | : : | : : : | : : |
| Db | 131 | SAPSQDGKTEAPRGAWKILGLFYAAALYPLAACAC-TAGHTAAHLLGSTLSWAHLGVQVW | 189 | | | | |
| QY | 42 | ----- | PCIPKSGYSSVVCVNATYCDSEDPPTFPALGTFRSYESTSRGRRMELSMGPIQ | 96 | | | |
| | | : : : | : : : | : | : : | : | : : |
| Db | 190 | QRAECPQVPKIYKYYSLLA----- | ----- | SLPLLLGLGFLSLWY----- | ----- | PVQ | 226 |
| QY | 97 | ----- | ANHTGTG | LLLLTLOPEQKFQKVKGFGGAMTDA | AAALNIALSPPAQNLLLSYFSEE | 151 | |
| | | : | : : | : | : | : : | : |
| Db | 227 | LVRFSRRTGAG----- | ----- | SKGLQSSYSEEYLRNLL----- | ----- | CRKKLGSSYHTSK | 268 |
| QY | 152 | GIGYNIIRVPMASCDFSIRTYTADTPDDFQLHNFSLP | PEEDTKIKIPLIHRALQLAQRPV | 211 | | | |
| | | : | : : | : | : : | : : | : |
| Db | 269 | HGFLSWARVCLRHCIY----- | ----- | TPQP----- | GFHLP----- | LKLVLSATLTGTAIYQV | 310 |

; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089905
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090691
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/095302
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095318
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095916
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/096146
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 60/096791
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 60/097986
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/098544
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099811
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100038
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101786
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101922
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/106178
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/106248
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/106464
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/106905
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/108787
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/108801
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/108849
; PRIOR FILING DATE: 1998-11-18
; PRIOR APPLICATION NUMBER: 60/112422
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113296
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119549
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/123618
; PRIOR FILING DATE: 1999-03-10
; PRIOR APPLICATION NUMBER: 60/125259
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 60/125775
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/126773
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: 60/127887
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/130232
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131022
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131445
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: 60/134287
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/140650
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/140723
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/141037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/144758
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/145698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/146222

```

; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 60/146963
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/149320
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149638
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/151733
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/164418
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/166361
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/169445
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169835

Query Match      2.9%  Score 84.5; DB 9; Length 667;
Best Local Similarity 19.0%; Pred. No. 14;
Matches 117; Conservative 65; Mismatches 202; Indels 233; Gaps 31;

QY 4 SSPSRE---ECPK-----PLSRVSIMAGSLTGLLLQAVSWAS-GAR-- 41
Db 131 SAPSQDKTEAPRGAWKILGLFYAALYPLAACA-TAGHTAAHLGLSTLSWAHLGVQVW 189

QY 42 -----PCIPKSPGYSSVVCNATYCDSDPPTFPALGTFSRYESTSRGRMELSMGPIQ 96
Db 190 QRAECPQVPKIKYYSLLA-----SLPLLGLGFLSLWY-----PVQ 226

QY 97 -----ANHTGTGLLTLQPEQKQKVGKGGAMTDAALNILALSPPAQNLLKSYFSEE 151
Db 227 LVRSFSRRTGAG-----SKGLQSSYSEYLRNLL-----CRKKLGSSYHTSK 268

QY 152 GIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRPV 211
Db 269 HGFLSWARVCLRHCIY-----TPQP-----GFHLP-----LKLVLSTLTGTALYQV 310

QY 212 SLASPTWPTWLKTNQAVN-----GKSLKQPCPD-----IYHQTWA----- 249
Db 311 ALLLVGVWPTIQVRAGVTTDVSYLLAGFIVLSEDKQEVVELVKHHLWALEVCYISAL 370

QY 250 -----RYFVKFLDAYAEHK-----LQFWAVTAENEPSAGLL-----SGY--PF 285
Db 371 VLSCLLTFLVLMRSLVTHRTNLRALHRGAALDLSPLHRSPHPSRQAIFCWMFSAYQTAF 430

QY 286 QCLGFTPEHQDFTARD-----LGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 431 ICLGLLVQOIIFFLGTALAFVLMPLV-----HGRNLLFRSLESSWPFWLTAL----- 481

QY 340 AAKYVHGIADVWYLDLAPAKATLGETHRLFPNT-----MLFAEACVGSKFWE 388
Db 482 -AVILQNMMAHWV--FL-----ETHDGHQPQLTNRRLVLYAATFLPLNLVGA----- 526

QY 389 QSVRLGSDRGMOYSHSIITNLLYHVVGWTDWNALNPEGGPNVVRNFVDSPIIVDITKD 448
Db 527 ---MVATW-----RVLLSALYNAIHLGQMDLSLLPPRA-----ATLD 560

QY 449 TFYKQPMFYHLGHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSVAVVVLNRSKDV 508
Db 561 -----PGYTYTYRNLKI-----EVSQSHPAMTAFCSLLQLAQSLLP 596

QY 509 LTIKDPVAVGFLFTISPG 525
Db 597 RTMAAPQ-----DSLRRPG 609
```

RESULT 15
US-10-106-092-4
; Sequence 4, Application US/10106092
; Patent No. US2002015604A1
; GENERAL INFORMATION:

```

; APPLICANT: Tufaro, Frank
; TITLE OF INVENTION: USE OF EXT GENES FOR THE TREATMENT OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/106,092
; FILING DATE: 25-Mar-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 920041.414C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-106-092-4

Query Match      2.9%  Score 84.5; DB 9; Length 728;
Best Local Similarity 20.8%; Pred. No. 16;
Matches 65; Conservative 42; Mismatches 97; Indels 109; Gaps 16;

QY 243 IYHQTWARYFVKFLDAYAEHKLQFW--AVTAENE-----PSAG 278
Db 24 IYYIT--LFSIVLLGLIATGMFQFWPHSHSISSNDWNVEKRSIRDVPVRLPADSPIERG 81

QY 279 LLGYPFQCL-----GFTPEHQ-RDFI-----ARDLGPPLAN--STHNVRLMLDDQ 324
Db 82 DLSCRMHTCFDVRGCFNPKNKIKYIYALKKYVDDFGVSVSNTISREYNELMAISDS- 140

QY 325 LLLPHWAKVVLTDPEAAKYVHGIADVWYLDLAPAKATLGETHRLFPNTMLFASEACVGS 384
Db 141 -----DYTDDINRACLFVPSIDVLNQNTLRIKETA----- 171

QY 385 KFEWQSVRLGSDRGMOYSHSIITNLLYHVV--GWTDNIALN-PE-----GGPNW- 432
Db 172 ---QAMAQLSRWDRGTNH-----LLENMLPGPPDYNTALDVPDRALLAGGFGFTWT 221

QY 433 -----VRNFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPEGSRQVGLVASQKNDLDAV 486
Db 222 YRQGYDVSIPIVYSPLSAEVDLPEKGPGRQYFL-----LSSQVGLHPEYREDLEAL 272

QY 487 ALMHPDGSVAVV 499
Db 273 QVKH--GESVLVL 283
```

Search completed: February 20, 2003, 11:24:18
Job time : 19 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: February 20, 2003, 11:08:31 ; Search time 14 Seconds
(without alignments)
1587.951 Million cell updates/sec
Title: US-10-024-197-25
Perfect score: 2865
Sequence: 1 MEFFSSPSRECKPLSRVSI.....GFLETISPGYSIHTYLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | | Description | |
|------------|--------|---------------|--------|----|-------------|--------------------|
| Result No. | Score | Query Match % | Length | ID | Description | |
| 1 | 2860 | 99.8 | 536 | 1 | GLCM_HUMAN | P04062 homo sapien |
| 2 | 2416.5 | 84.3 | 515 | 1 | GLCM_MOUSE | P17439 mus musculu |
| 3 | 123.5 | 4.3 | 562 | 1 | EXG2_YEAST | P52911 saccharomyc |
| 4 | 105 | 3.7 | 2700 | 1 | ZAN_HUMAN | Q9Y493 homo sapien |
| 5 | 104.5 | 3.6 | 1738 | 1 | CO4_MOUSE | P01029 mus musculu |
| 6 | 99.5 | 3.5 | 466 | 1 | GUN5_THEFU | Q01786 thermomonos |
| 7 | 99 | 3.5 | 668 | 1 | GALC_MOUSE | P54818 mus musculu |
| 8 | 97.5 | 3.4 | 669 | 1 | GALC_MACMU | O02791 macaca mula |
| 9 | 97 | 3.4 | 641 | 1 | IMD_ARTGO | O44052 arthrobaete |
| 10 | 96 | 3.4 | 1711 | 1 | PTPO_RAT | Q64612 rattus norv |
| 11 | 95 | 3.3 | 512 | 1 | HMZ2_ARATH | O04921 arabidopsis |
| 12 | 95 | 3.3 | 514 | 1 | C11A_DASAM | Q92045 dasyatis am |
| 13 | 94.5 | 3.3 | 669 | 1 | GALC_HUMAN | P54803 homo sapien |
| 14 | 94.5 | 3.3 | 822 | 1 | PHK_LACLA | O9cfh4 lactococcus |
| 15 | 92.5 | 3.2 | 6486 | 1 | TYCC_BACBR | O30409 b tyrocidin |
| 16 | 92 | 3.2 | 774 | 1 | FECA_ECOLI | P13036 escherichia |
| 17 | 91 | 3.2 | 1774 | 1 | MSAS_PENPA | P22367 penicillium |
| 18 | 89.5 | 3.1 | 1748 | 1 | POLR_ELV | P35928 erysimum la |
| 19 | 89.5 | 3.1 | 2025 | 1 | TTC3_HUMAN | P53804 homo sapien |
| 20 | 89 | 3.1 | 555 | 1 | HUTU_XANCP | P58988 xanthomonas |
| 21 | 88.5 | 3.1 | 350 | 1 | DCUP_SYNY3 | P54224 synechocyst |
| 22 | 88.5 | 3.1 | 464 | 1 | YRG5_CAEEL | Q09406 caenorhabdi |
| 23 | 88.5 | 3.1 | 4303 | 1 | PKD1_HUMAN | P98161 homo sapien |
| 24 | 88 | 3.1 | 556 | 1 | PDPK_HUMAN | O15530 homo sapien |
| 25 | 88 | 3.1 | 565 | 1 | SAPA_HAEIN | P45285 haemophilus |
| 26 | 87.5 | 3.1 | 874 | 1 | SYLM_SCHPO | Q09828 schizosacch |
| 27 | 87.5 | 3.1 | 920 | 1 | AKH1_MAIZE | P49079 zea mays (m |
| 28 | 87 | 3.0 | 391 | 1 | PCL_ECTHA | P42516 ectothiorho |
| 29 | 87 | 3.0 | 555 | 1 | HUTU_XANAC | P58987 xanthomonas |
| 30 | 87 | 3.0 | 627 | 1 | CACP_COLLI | P52826 columba liv |
| 31 | 87 | 3.0 | 1274 | 1 | SRF3_BACSU | Q08787 bacillus su |
| 32 | 86.5 | 3.0 | 600 | 1 | SPO8_YEAST | P41833 saccharomyc |
| 33 | 86.5 | 3.0 | 1448 | 1 | CPSA_ARATH | Q9fgr0 arabidopsis |

ALIGNMENTS

| RESULT 1 | | | | | ALIGNMENTS | |
|----------|--|-----------|------|---------|------------|--|
| ID | GLCM_HUMAN | STANDARD; | PRT; | 536 AA. | | |
| AC | P04062; | | | | | |
| DT | 01-NOV-1986 (Rel. 03, Created) | | | | | |
| DT | 01-AUG-1990 (Rel. 15, Last sequence update) | | | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | | | |
| DE | Glucosylceramidase precursor (EC 3.2.1.45) (Beta-glucocerebrosidase) | | | | | |
| DE | (Acid beta-glucosidase) (D-glucosyl-N-acylsphingosine glucosylhydrolase) | | | | | |
| DE | (Alglucerase) (Imiglucerase). | | | | | |
| GN | GBA OR GC. | | | | | |
| OS | Homo sapiens (Human). | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| OX | NCBI_TaxID=9606; | | | | | |
| RN | [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | TISSUE=Placenta; | | | | | |
| RX | MEDLINE=86042651; PubMed=3864160; | | | | | |
| RA | Sorge J., West C., Westwood B., Beutler E.; | | | | | |
| RT | "Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA."; | | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 82:7289-7293(1985). | | | | | |
| RN | [2] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | TISSUE=Hepatoma; | | | | | |
| RX | MEDLINE=86085859; PubMed=3001061; | | | | | |
| RA | Tsuji S., Choudary P.V., Martin B.M., Winfield S., Barranger J.A.; | | | | | |
| RT | Guinn S.I.; | | | | | |
| RT | "Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase."; | | | | | |
| RL | J. Biol. Chem. 261:50-53(1986). | | | | | |
| RN | [3] | | | | | |
| RP | SEQUENCE OF 1-11 FROM N.A. | | | | | |
| RX | MEDLINE=88195776; PubMed=3359914; | | | | | |
| RA | Reiner O., Wigderson M., Horowitz M.; | | | | | |
| RT | "Structural analysis of the human glucocerebrosidase genes."; | | | | | |
| RL | DNA 7:107-116(1988). | | | | | |
| RN | [4] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RA | Horowitz M.; | | | | | |
| RL | Submitted (MAY-1988) to the EMBL/GenBank/DBJ databases. | | | | | |
| RN | [5] | | | | | |
| RP | SEQUENCE OF 1-45 FROM N.A. | | | | | |
| RX | MEDLINE=88074307; PubMed=3687939; | | | | | |
| RA | Sorge J.A., West C., Kuhl W., Treger L., Beutler E.; | | | | | |
| RT | "The human glucocerebrosidase gene has two functional ATG initiator codons."; | | | | | |
| RL | Am. J. Hum. Genet. 41:1016-1024(1987). | | | | | |
| RN | [6] | | | | | |
| RP | SEQUENCE OF 40-44. | | | | | |
| RC | TISSUE=Placenta; | | | | | |
| RA | Martin B.M., Murray G.J., Coligan J.E., Raum M., Brady R.O.; | | | | | |
| RA | Barranger J.A.; | | | | | |
| RT | "Structural studies of human placental glucocerebrosidase."; | | | | | |
| RL | Fed. Proc. 43:1869-1869(1984). | | | | | |

Q13049 homo sapien
Q03155 escherichia
P29336 streptococc
P51853 pseudomonas
P40724 salmonella
Q61085 mus musculu
P52087 pseudomonas
P32823 alteromonas
P75189 mycoplasma
O44783 drosophila
Q93063 homo sapien
Q10639 mycobacteri

RN [7] SEQUENCE OF 469-520.
RP TISSUE=Placenta;
RX MEDLINE=86149363; PubMed=3456607;
RA Dinur T., Osiecki K.M., Legler G., Gatt S., Desnick R.J.,
RA Grabowski G.A.;
RT "Human acid beta-glucosidase: Isolation and amino acid sequence of a
peptide containing the catalytic site.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1660-1664(1986).
RN [8]
RP REVIEW ON GD VARIANTS.
RX MEDLINE=94163182; PubMed=8118460;
RA Horowitz M., Zimran A.;
RT "Mutations causing Gaucher disease.";
RL Hum. Mutat. 3:1-11(1994).
RN [9]
RP REVIEW ON GD VARIANTS.
RX MEDLINE=97044507; PubMed=8889578;
RA Beutler E., Gelbart T.;
RT "Glucocerebrosidase (Gaucher disease).";
RL Hum. Mutat. 8:207-213(1996).
RN [10]
RP REVIEW ON GD VARIANTS.
RX MEDLINE=99459154; PubMed=10527671;
RA Tayebi N., Stone D.L., Sidransky E.;
RT "Type 2 gaucher disease: an expanding phenotype.";
RL Mol. Genet. Metab. 68:209-219(1999).
RN [11]
RP REVIEW ON GD VARIANTS.
RX MEDLINE=20115339; PubMed=10649495;
RA Stone D.L., Tayebi N., Orvisky E., Stubblefield B., Madike V.,
RA Sidransky E.;
RT "Glucocerebrosidase gene mutations in patients with type 2 Gaucher
disease.";
RL Hum. Mutat. 15:181-188(2000).
RN [12]
RP VARIANT GD TYR-255.
RX MEDLINE=90343254; PubMed=1974409;
RA Beutler E., Gelbart T.;
RT "Gaucher disease associated with a unique KpnI restriction site:
identification of the amino-acid substitution.";
RL Ann. Hum. Genet. 54:149-153(1990).
RN [13]
RP VARIANTS GD.
RX MEDLINE=90274868; PubMed=1972019;
RA Hong C.M., Ohashi T., Yu X.J., Weiler S., Barranger J.A.;
RT "Sequence of two alleles responsible for Gaucher disease.";
RL DNA Cell Biol. 9:233-241(1990).
RN [14]
RP VARIANTS GD.
RX MEDLINE=93162655; PubMed=8432537;
RA Beutler E., Gelbart T., West C.;
RT "Identification of six new Gaucher disease mutations.";
RL Genomics 15:203-205(1993).
RN [15]
RP VARIANT GD HIS-535.
RX MEDLINE=94379086; PubMed=7916532;
RA Choy F.Y.M., Wei C., Applegarth D.A., McGillivray B.C.;
RT "DNA analysis of an uncommon missense mutation in a Gaucher disease
patient of Jewish-Polish-Russian descent.";
RL Am. J. Med. Genet. 51:156-160(1994).
RN [16]
RP VARIANT GD ASN-438.
RX MEDLINE=94156356; PubMed=8112750;
RA Beutler E., Gelbart T.;
RT "Two new Gaucher disease mutations.";
RL Hum. Genet. 93:209-210(1994).
RN [17]
RP VARIANTS GD SER-409 AND CYS-457.
RX MEDLINE=94357595; PubMed=8076951;
RA Tuteja R., Tuteja N., Lilliu F., Bembli B., Galanello R., Cao A.,
RA Baralle F.E.;
RT "Y418C: a novel mutation in exon 9 of the glucocerebrosidase gene of

RT a patient with Gaucher disease creates a new Bgl I site.";
RL Hum. Genet. 94:314-315(1994).
RN [18]
RP VARIANTS GD D-215; T-221; R-241; Q-296; C-324; G-417 AND N-419.
RX MEDLINE=96382740; PubMed=8790604;
RA Beutler E., Demina A., Gelbart T.;
RT "Glucocerebrosidase mutations in Gaucher disease.";
RL Mol. Med. 1:82-90(1994).
RN [19]
RP VARIANTS GD SER-409; HIS-448; PRO-483 AND CYS-502.
RX MEDLINE=95353281; PubMed=7627184;
RA Cormand B., Vilageliu L., Burguera J.M., Balcells S.,
RA Gonzalez-Duarte R., Grinberg D., Chabas A.;
RT "Gaucher disease in Spanish patients: analysis of eight mutations.";
RL Hum. Mutat. 5:303-309(1995).
RN [20]
RP VARIANT GD SER-217.
RX MEDLINE=95353289; PubMed=7627192;
RA Choy F.Y.M., Wei C.;
RT "Identification of a new mutation (P178S) in an African-American
patient with type 2 Gaucher disease.";
RL Hum. Mutat. 5:345-347(1995).
RN [21]
RP VARIANTS GD SER-409; LEU-426; LEU-433 AND PRO-483.
RX MEDLINE=97092100; PubMed=8937765;
RA Morar B., Lane A.B.;
RT "The molecular characterization of Gaucher disease in South Africa.";
RL Clin. Genet. 50:78-84(1996).
RN [22]
RP VARIANTS GD LEU-54; GLU-85 AND SER-227.
RX MEDLINE=96271984; PubMed=8829654;
RA Kim J.-W., Liou B.B., Lai M.-Y., Ponce E., Grabowski G.A.;
RT "Gaucher disease: identification of three new mutations in the Korean
and Chinese (Taiwanese) populations.";
RL Hum. Mutat. 7:214-218(1996).
RN [23]
RP VARIANTS GD HIS-352 AND GLN-398.
RX MEDLINE=96271993; PubMed=8829663;
RA Cormand B., Vilageliu L., Balcells S., Gonzalez-Duarte R., Chabas A.,
RA Grinberg D.;
RT "Two novel (1098insA and Y313H) and one rare (R359Q) mutations
detected in exon 8 of the beta-glucocerebrosidase gene in Gaucher's
disease patients.";
RL Hum. Mutat. 7:272-274(1996).
RN [24]
RP VARIANT GD THR-435.
RX MEDLINE=97044520; PubMed=8889591;
RA Amaral O., Pinto E., Fortuna M., Lacerda L., Sa Miranda M.C.;
RT "Type 1 Gaucher disease: identification of N396T and prevalence of
glucocerebrosidase mutations in the Portuguese.";
RL Hum. Mutat. 8:280-281(1996).
RN [25]
RP VARIANTS GD LEU-437 AND ILE-530.
RX MEDLINE=96373808; PubMed=8780099;
RA Seeman P.J.V., Finckh U., Hoepfner J., Lakner V., Liebisch I.,
RA Grau G., Rolfs A.;
RT "Two new missense mutations in a non-Jewish Caucasian family with type
3 Gaucher disease.";
RL Neurology 46:1102-1107(1996).
RN [26]
RP VARIANTS GD LEU-414 AND THR-441.
RX MEDLINE=97325776; PubMed=9182788;
RA Cormand B., Grinberg D., Gort L., Fiumara A., Barone R., Vilageliu L.,
RA Chabas A.;
RT "Two new mild homozygous mutations in Gaucher disease patients:
clinical signs and biochemical analyses.";
RL Am. J. Med. Genet. 70:437-443(1997).
RN [27]
RP VARIANTS GD V-76; E-85; W-87; S-227; I-252 AND P-483.
RX MEDLINE=97360189; PubMed=9217217;
RA Choy F.Y.M., Humphries M.L., Shi H.;
RT "Identification of two novel and four uncommon missense mutations
among Chinese Gaucher disease patients.";

Query Match 99.8%; Score 2860; DB 1; Length 536;
Best Local Similarity 99.8%; Pred. No. 2.7e-233;
Matches 535; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEFSSPSRECKPLSRVIMAGSLTGLLLQAVSWAGARPCIPKSGYSSVVCVNCAT 60
Db 1 MEFSSPSRECKPLSRVIMAGSLTGLLLQAVSWAGARPCIPKSGYSSVVCVNCAT 60

QY 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTQPEQKFKVKG 120
Db 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTQPEQKFKVKG 120

QY 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180
Db 121 GGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDD 180

QY 181 FQLHNSLPEEDTKLTIPLIHRALQLAQRPVSLASPTWLTNGAVNGKSLKGQP 240
Db 181 FQLHNSLPEEDTKLTIPLIHRALQLAQRPVSLASPTWLTNGAVNGKSLKGQP 240

QY 241 GDIYHQWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHORDFIA 300
Db 241 GDIYHQWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHORDFIA 300

QY 301 RDLGPTLANSTHNNVRLMLDDQRLPLHWAQVLTDPBAKYVHGIHVWYLDLAPAK 360
Db 301 RDLGPTLANSTHNNVRLMLDDQRLPLHWAQVLTDPBAKYVHGIHVWYLDLAPAK 360

QY 361 ATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDWDRGMQYSHSIITNLLYHVWGTDW 420
Db 361 ATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDWDRGMQYSHSIITNLLYHVWGTDW 420

QY 421 NLALNPEGGPNVWVNFVDSPIIVDITKDFYKQPMFYHLGHFSKFIPEGQRVGLVASQK 480
Db 421 NLALNPEGGPNVWVNFVDSPIIVDITKDFYKQPMFYHLGHFSKFIPEGQRVGLVASQK 480

QY 481 NDLDVALMHDPDGSVAVVVLNRRSSKDVPLTIKDPVAGVLETISPYSIHTYLWRRQ 536
Db 481 NDLDVALMHDPDGSVAVVVLNRRSSKDVPLTIKDPVAGVLETISPYSIHTYLWRRQ 536

RESULT 2
GLCM_MOUSE STANDARD; PRT; 515 AA.

AC P17439;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Glucosylceramidase precursor (EC 3.2.1.45) (Beta-glucocerebrosidase)
DE (Acid beta-glucosidase) (D-glucosyl-N-acylsphingosine glucosylhydrolase).
GN GBA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89296941; PubMed=2740343;
RA O'Neill R.R., Tokoru T., Kozak C.A., Brady R.O.;
RT "Comparison of the chromosomal localization of murine and human
RT glucocerebrosidase genes and of the deduced amino acid sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5049-5053(1989).
RN [2]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=92272748; PubMed=1317175;
RA Carstea E.D., Murray G.J., O'Neill R.R.;
RT "Molecular and functional characterization of the murine
RT glucocerebrosidase gene";
RL Biochem. Biophys. Res. Commun. 184:1477-1483(1992).
CC -1- CATALYTIC ACTIVITY: D-glucosyl-N-acylsphingosine + H(2)O - D-
CC glucose + N-acylsphingosine.

CC -1- SUBCELLULAR LOCATION: LYSOSOMAL, MEMBRANE BOUND.
CC -1- SIMILARITY: BELONGS TO FAMILY 30 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24119; AAA37671.1; -.
DR EMBL; M89949; AAA37665.1; -.
DR PIR; A32931; A32931.
DR MGD; MGI:95665; Gba.
DR InterPro; IPR001139; GH_30.
DR Pfam; PF02055; Glyco_hydro_30; 1.
DR PRINTS; PR00843; GLHYDRLASE30.
KW Hydrolase; Glycosidase; Sphingolipid metabolism; Glycoprotein;
KW Lysosome; Membrane; Gaucher disease; Signal.
FT SIGNAL 1 19
FT CHAIN 20 515 GLUCOSYLCERAMIDASE.
FT ACT_SITE 254 254 PROTON DONOR (POTENTIAL).
FT ACT_SITE 358 358 NUCLEOPHILE (BY SIMILARITY).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 515 AA; 57621 MW; 7CD9176085FE2CB CRC64;

Query Match 84.3%; Score 2416.5; DB 1; Length 515;
Best Local Similarity 86.4%; Pred. No. 6.1e-196;
Matches 446; Conservative 30; Mismatches 39; Indels 1; Gaps 1;

QY 21 MAGSLTGLLLQAVSWAGARPCIPKSGYSSVVCVNCATYCDSPPTFPALGTFSTRYE 80
Db 1 MAARLIGFLFQAVSWAYGAQPCIPKSGYSSVVCVNCATYCDSPPTFPALGTFSTRYE 60

QY 81 STRSGRRMELSMGPIQANHTGTGLLLTQPEQKFKVKGFGAMTDAALNIALSPPAQ 140
Db 61 STRSGRRMELSVGAIQANRTGTGLLLTQPEKFKVKGFGAMTDAALNIALSPPTQ 120

QY 141 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYADTPDDFQLHNSLPEEDTKLKIPLI 200
Db 121 KLLRSYFSTNGIEYNIIRVPMASCDFSIRTYADTPDNPQLSNFSLPEEDTKLKIPLI 180

QY 201 HRALQLAQRPSVLLASPTWLTWKTNGAVNGKSLKGQGDYIYHQWARYFVKFLDAYA 260
Db 181 HQALKMSSRPISLFASTWLTWKTNGRNGKSLKGQGDYIYHQWARYFVKFLDAYA 240

QY 261 EHKLQFWAVTAENEPSAGLLSGYFPQCLGFTPEHORDFIARDLGPPTLANSTHNNVRLML 320
Db 241 KYGLRFWAVTAENEPATAGLTGTGYPFQCLGFTPEHORDFISRDLPALANSS-HDVKLLML 299

QY 321 DDQRLLLPHWAKVVLTDPEAAKYVHGIHVWYLDLAPAKATLGETHRLFPNTMLFASEA 380
Db 300 DDQRLLLPRWAEVLSLSDPEAAKYVHGIHVWYLDLAPAKATLGETHRLFPNTMLFASEA 359

QY 381 CVGSKFWEQSVRLGSDWDRGMQYSHSIITNLLYHVWGTWDLNLPNPEGPNVWVNFVDS 440
Db 360 CVGSKFWEQSVRLGSDWDRGMQYSHSIITNLLYHVWGTWDLNLPNPEGPNVWVNFVDS 419

QY 441 IIVDITKDFYKQPMFYHLGHFSKFIPEGQRVGLVASQKNDLDAVALMHDPDGSVAVVVL 500
Db 420 IIVDIPKDAFYKQPMFYHLGHFSKFIPEGQRVGLVASQKNDLDAVALMHDPDGSVAVVVL 479

QY 501 NRSSKDVPLTIKDPVAGVLETISPYSIHTYLWRRQ 536
Db 480 NRSSKDVPLTIKDPVAGVLETISPYSIHTYLWRRQ 515

CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE

CC ZONA PELLUCIDA.

CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,

CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF

CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.

CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT

CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).

CC -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

CC -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF053356; AAC78790.1; -.

DR EMBL; U83191; AAC51208.1; -.

DR Genew; HGNC:12857; ZAN.

DR MIM; 602372; -.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000998; MAM_domain.

DR InterPro; IPR002919; TIL_Cysrich.

DR InterPro; IPR003328; TILa_Cysrich.

DR InterPro; IPR001846; VWF_D.

DR Pfam; PF00094; vwd; 4.

DR Pfam; PF00629; MAM; 4.

DR Pfam; PF01826; TIL; 5.

DR Pfam; PF02345; TILa; 4.

DR SMART; SM00216; VWD; 1.

DR PROSITE; PS01186; EGF_2; 3.

DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS50060; MAM_2; 4.

DR Glycoprotein; Transmembrane; Cell adhesion; Repeat.

FT NON_TER 1 1

FT DOMAIN <1 109 MAM 1.

FT DOMAIN 112 136 MAM 2 (PARTIAL).

FT DOMAIN 161 326 MAM 3.

FT DOMAIN 322 446 MAM 4.

FT DOMAIN 483 951

FT DOMAIN 953 1065

FT DOMAIN 1066 1454

FT DOMAIN 1455 1861

FT DOMAIN 1862 2292

FT DOMAIN 2293 2684

FT DOMAIN ? 74

FT CARBOHYD 74 74

FT CARBOHYD 403 403

FT CARBOHYD 1023 1023

FT CARBOHYD 1099 1099

FT CARBOHYD 1618 1618

FT CARBOHYD 1737 1737

FT CARBOHYD 1832 1832

FT CARBOHYD 1878 1878

FT CARBOHYD 2136 2136

FT CARBOHYD 2505 2505

FT CONFLICT 2374 2379

FT NON_TER 2700 2700

SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match 3.7%; Score 105; DB 1; Length 2700;

Best Local Similarity 20.3%; Pred. No. 4.5;

Matches 125; Conservative 55; Mismatches 174; Indels 262; Gaps 35;

QY 29 LLLQAVSWAGARPCIPKSF-----YSSVVCVCNATYCDSDPPTFPALG 74

Db :|||:|||||:|

15 VLLSPVSLSSG---CLSFSEFHYILRGQSPGALHIYASV-----LG 52

QY 75 TFSRYESTRSGRRMELSMGP-----IQANHGTGTLTLLQPEQKFKQVKFGGAMTDAAL 130

Db :|||:|||||:|

53 SI-RKHTLESGQ-----PGENWQAVSVNYTAVGRI-----QFVAVGVFGKTPPEAVAV 99

QY 131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDQLHNFSLPE 190

Db :|||:|||||:|

100 DATSIAPCGEG-----FPQCDFFEDNAHPFCD----- 125

QY 191 EDTKLKLIPLIHRALQLAQRVSLASPWTSPTWLKT-----NGAVNGKGSCLKG 238

Db :|||:|||||:|

126 -----WVQTSGDGCHWALGHKNGPVHGMGPAGG 153

QY 239 QP-GDIYHQ-----TWARYFVKFLDAYAEHKLQFWAVTAENEP--SAGLLSGYP 284

Db :|||:|||||:|

154 FPNADVLTAQDFEDDAKPLCDWSQVSADDED-----WVRASGPSPTGSTGAPGGYP 204

QY 285 -----FQCLGFTPEHQRDFIAR-----DLGPTLANSTHH-----NVRLLMLD 321

Db :|||:|||||:|

205 NGEGLHMEESNSF-----HRGGVARLLSPDLWEQGPLCVHFAHMFGLSWGQAQLRLLLS 260

QY 322 DQR-----LLLPHWAKVVLTDPEAAKYVHGIYAVHWHYLDL-APAKATLGTETRLFPNTML 375

Db :|||:|||||:|

261 GEEGRRPDVWLKHW-----NTQRPSSWMLTTVTVPAGFTL-PTRLMFEGR- 304

QY 376 FASEACVGSKFWEQSVRLGSDWRGMQY-----SHS-----IITNLL 411

Db :|||:|||||:|

305 -GSTAYLDIALDALSIIRGSCNRGGHYIYLEADEFSHAGQSVRLVSRPFCAPDICVEFA 363

QY 412 YHVVG---WTDWNLANPEGG-----PNWVRNFVDSPI---IVDITKDTFYKQPM---FYHL 459

Db :|||:|||||:|

364 YHMYGLGEGTMLELLGSPAGSPPIPLMKRVGSRQPYQWTSVTVPSGHQPMQLIFKGI 423

QY 460 GHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSAAVVVVLNRSSKDVPLTIKDP- -VG 517

Db :|||:|||||:|

424 -----QGSNTASVVAM-----GFTLINP-GTCPVKVLPPELPPVSPVSSSTGPSETTG 468

QY 518 FLE--TIS---PGYSI 528

Db :|||:|||||:|

469 LTENPTISTKKPTVSI 484

RESULT 5

CO4_MOUSE

ID CO4_MOUSE STANDARD; PRT; 1738 AA.

AC P01029; Q61859;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Complement C4 precursor [Contains: C4A anaphylatoxin].

GN C4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85298264; PubMed=3862104;

RA Sepich D.S., Noonan D.J., Ogata R.T.;

RT "Complete cDNA sequence of the fourth component of murine

RT complement.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:5895-5899(1985).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=B12.WR;

RX MEDLINE=87309760; PubMed=3624865;

RA Rosa P.A., Sepich D.S., Robins D.M., Ogata R.T.;

RT "Constitutive expression of Slp genes in mouse strain B10.WR directed

RT by C4 regulatory sequences.";

RL J. Immunol. 139:1568-1577(1987).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=B12.WR; TISSUE=Liver;

RX MEDLINE=89380278; PubMed=2777798;

RA Ogata R.T., Rosa P.A., Zepf N.E.;
RT "Sequence of the gene for murine complement component C4.";
RL J. Biol. Chem. 264:16565-16572(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=FM; TISSUE=Liver;
RX MEDLINE=85289294; PubMed=2993295;
RA Nonaka M., Nakayama K., Yeul Y.D., Takahashi M.;
RT "Complete nucleotide and derived amino acid sequences of the fourth
component of mouse complement (C4). Evolutionary aspects.";
RL J. Biol. Chem. 260:10936-10943(1985).
RN [5]
RP SEQUENCE OF 651-810 AND 924-1083 FROM N.A.
RX MEDLINE=85038607; PubMed=6208559;
RA Nonaka M., Takahashi M., Natsuume-Sakai S., Nonaka M., Tanaka S.,
RA Shimizu A., Honjo T.;
RT "Isolation of cDNA clones specifying the fourth component of mouse
complement and its isotype, sex-limited protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:6822-6826(1984).
RN [6]
RP SEQUENCE OF 1099-1142 FROM N.A.
RX STRAIN=B10.WR, C57BL/6, C3H/He, CBA/J, B10.BR, and DBA/2;
RX MEDLINE=90353398; PubMed=2387317;
RA Ogata R.T., Zepf N.E.;
RT "C4 from C4-high and C4-low mouse strains have identical sequences in
the region corresponding to the isotype-specific segment of human
C4.";
RL Eur. J. Immunol. 20:1607-1610(1990).
RN [7]
RP SEQUENCE OF 1105-1449 FROM N.A.
RX MEDLINE=85166208; PubMed=3856857;
RA Levi-Strauss M., Tosi M., Steinmetz M., Klein J., Meo T.;
RT "Multiple duplications of complement C4 gene correlate with H-2-
controlled testosterone-independent expression of its sex-limited
isoform, C4-Slp.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1746-1750(1985).
RN [8]
RP SEQUENCE OF 1257-1376 FROM N.A.
RX MEDLINE=85038859; PubMed=6149581;
RA Tosi M., Levi-Strauss M., Duponchel C., Meo T.;
RT "Sequence heterogeneity of murine complementary DNA clones related to
the C4 and C4-Slp isoforms of the fourth complement component.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:389-394(1984).
RN [9]
RP SEQUENCE OF 1360-1511 FROM N.A.
RX MEDLINE=83273751; PubMed=6192448;
RA Ogata R.T., Shreffler D.C., Sepich D.S., Lilly S.P.;
RT "cDNA clone spanning the alpha-gamma subunit junction in the
precursor of the murine fourth complement component (C4).";
RL Proc. Natl. Acad. Sci. U.S.A. 80:5061-5065(1983).
RN [10]
RP SEQUENCE OF 1-128 FROM N.A.
RC STRAIN=FM; TISSUE=Liver;
RX MEDLINE=86031969; PubMed=2997024;
RA Nonaka M., Nakayama K., Yeul Y.D., Shimizu A., Takahashi M.;
RT "Molecular cloning and characterization of complementary and genomic
DNA clones for mouse C4 and Slp.";
RL Immunol. Rev. 87:81-99(1985).
RN [11]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=87017050; PubMed=3464002;
RA Nonaka M., Kimura H., Yeul Y.D., Yokoyama S., Nakayama K.,
RA Takahashi M.;
RT "Identification of the 5'-flanking regulatory region responsible for
the difference in transcriptional control between mouse complement C4
and Slp genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7883-7887(1986).
CC -1- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESSED BY
ACTIVATED C1 WHICH REMOVE FROM THE ALPHA CHAIN THE C4A
ANAPHYLATOXIN.
CC -1- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
AND, PRIOR TO SECRETION, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER

CC OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
CC -1- MISCELLANEOUS: C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
CC PROTEIN.
CC -1- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K00019; AAA39554.1; -
CC EMBL: M11729; AAA39506.1; -
CC EMBL: M12968; AAA39558.1; -
CC EMBL: M12970; AAA39555.1; -
CC EMBL: M12972; AAA39556.1; -
CC EMBL: M12969; AAA39559.1; -
CC EMBL: M11789; AAA39557.1; -
CC EMBL: K02798; AAC42021.1; -
CC EMBL: M17440; AAA39561.1; -
CC EMBL: M14225; AAA39563.1; -
CC EMBL: M14226; AAA39684.1; -
CC EMBL: X55493; CAA39112.1; -
CC EMBL: X55495; CAA39114.1; -
CC PIR: A01264; A01264.
CC PIR: A21692; A21692.
CC PIR: A22039; A22039.
CC PIR: A24558; A24558.
CC PIR: A29059; A29059.
CC HSSP: P01031; 1KJS.
CC MGD: MGI:88228; C4.
CC InterPro: IPR002890; A2M_N.
CC InterPro: IPR000020; Anaphylatoxin.
CC InterPro: IPR001840; Anaphylatoxn.
CC InterPro: IPR001599; MacrogloblnA2.
CC InterPro: IPR001134; Netrin_C.
CC Pfam: PF00207; A2M; 1.
CC Pfam: PF01759; NTR; 1.
CC Pfam: PF01821; ANATO; 1.
CC Pfam: PF01835; A2M_N; 1.
CC PRINTS: PR00004; ANAPHYLATOXN.
CC ProDom: PD003264; Anaphylatoxin; 1.
CC SMART: SM00104; ANATO; 1.
CC PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; 1.
CC PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
CC PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
CC Complement pathway; Plasma; Glycoprotein; Sulfation; Signal;
CC Inflammatory response.
CC SIGNAL 1 19
CC CHAIN 20 673 COMPLEMENT C4, BETA CHAIN.
CC PROPEP 674 677 COMPLEMENT C4, ALPHA CHAIN.
CC CHAIN 678 1443 COMPLEMENT C4, GAMMA CHAIN.
CC PROPEP 1444 1447 C4A ANAPHYLATOXIN.
CC CHAIN 1448 1738 ANAPHYLATOXIN-LIKE.
CC PEPTIDE 678 753 BY SIMILARITY.
CC DOMAIN 700 734 BY SIMILARITY.
CC DISULFID 700 726 BY SIMILARITY.
CC DISULFID 701 733 BY SIMILARITY.
CC DISULFID 714 734 BY SIMILARITY.
CC THIOLEST 1006 1009 SULFATION.
CC MOD_RES 1413 1416 SULFATION.
CC MOD_RES 1416 1417 SULFATION.
CC MOD_RES 1417 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 224 224 N-LINKED (GLCNAC. . .).
CC CARBOHYD 743 743 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1387 1387 F -> Y (IN REF. 4).
CC CONFLICT 132 132 G -> E (IN REF. 4).
CC CONFLICT 327 327 Q -> E (IN REF. 4).
CC CONFLICT 570 570 R -> G (IN REF. 5).
CC CONFLICT 720 720

Db 106 ---TDGWIEGSSRTGNGYITSY-NDWQHDWA-YWANYL-----AARKMKLGYYNPLMVH 157
QY 270 TAENEPSAGLLSGYPFQCLG--FTPEHORDFIARDLGPTLANSTHNVRLMLDDQRLLL 327
Db 158 RAAVEDASKTVLGRPDVKIADLVVP---GDFEARDIGGN-----QLYWLD----- 199
QY 328 PHWAKVVLTPDAKYVHGIHVH-----YL--DFA-----PAKATLGETH-----RLF 370
Db 200 -----VTKSAKEYVQGVYRYFKDLGVPYLRIDFLSWYEDGRDANIGQVNAPHGRANY 252
QY 371 PNTMLFASEAC-----VGSKFWEQSVRL-----GSWDR--GMQYSHS 405
Db 253 ELALSWINEAAGEDMEVSLVMPHMFQDGSALANGDLVRLINADADKGGWDRLSGMRQN-- 310
QY 406 IITNLLYHVVGWTD-WNALNPEGGPNVVRNF-----PWNANPFCGFTGWSHRNGRGQLILGDGDMRASTFA 350
Db 311 -----WODAW-----PWNANPFCGFTGWSHRNGRGQLILGDGDMRASTFA 350
QY 437 -----VDSPI-----IVDTKDTFYKQPMFYHLGHF 462
Db 351 SDEERKTMNLMVAAGSPLAIADTYQIGNNAWVYTNKEVQLNADGLVKGKPLYSATPF 410
QY 463 SKFIPEGSRQVGLVASQKNDLDAVALMHPDGSAAVVVVLNRSSKD 506
Db 411 SK--DPGSR-----DTERWAGQLPDGSGWGVALFNRSSTE 442

RESULT 10

PTPO_RAT ID PTPO_RAT STANDARD; PRT; 1711 AA.
AC Q64612;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Osteotesticular protein tyrosine phosphatase precursor (EC 3.1.3.48) (OST-PTP).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus. NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Osteosarcoma;
RX MEDLINE-95074080; PubMed=7527035;
RA Mauro L.J., Olmsted E.A., Skrobacz B.M., Mourey R.J., Davis A.R., Dixon J.E.;
RA "Identification of a hormonally regulated protein tyrosine phosphatase associated with bone and testicular differentiation."; J. Biol. Chem. 269:30659-30667(1994).
CC -!- FUNCTION: MAY FUNCTION IN SIGNALING PATHWAYS DURING BONE REMODELING, AS WELL AS SERVE A BROADER ROLE IN CELL INTERACTIONS ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS. OPTIMAL PH FOR PHOSPHATASE ACTIVITY IS 5.6.
CC ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing. A presumed alternate transcript of 4.8-5.0 kilobases, which may lack PTP domains, is present in proliferating osteoblasts, but not detectable at other stages.
CC -!- TISSUE SPECIFICITY: BONE AND TESTIS. IN THE LATTER, RESTRICTED TO THE BASAL PORTION OF THE SEMINIFEROUS TUBULE.
CC -!- DEVELOPMENTAL STAGE: UP-REGULATED IN DIFFERENTIATING CULTURES OF PRIMARY OSTEOBLASTS AND DOWN-REGULATED IN LATE STAGE MINERALIZING CULTURES. IN TESTIS, EXPRESSION IS HIGHEST BETWEEN STAGES I AND VII WHEN MATURING SPERMATIDS REMAIN BURIED WITHIN THE SERTOLI EPITHELIUM.
CC -!- INDUCTION: BY PARATHYROID HORMONE AND CYCLIC AMP ANALOGS.
CC -!- PFM: THE CYTOPLASMIC DOMAIN CONTAINS POTENTIAL PHOSPHORYLATION SITES.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -!- SIMILARITY: CONTAINS 10 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC EMBL; L36884; AAA63911.1; --
DR HSSP; P18052; IYFO.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00041; fn3; 7.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00014; ENTPEIIL.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00194; PTPc; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolyase; Transmembrane; Repeat; Signal; Glycoprotein.

FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1711 OSTEOTESTICULAR PROTEIN TYROSINE PHOSPHATASE.
FT DOMAIN 18 1074 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1075 1095 POTENTIAL.
FT DOMAIN 1096 1711 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 32 124 FIBRONECTIN TYPE-III 1.
FT DOMAIN 125 215 FIBRONECTIN TYPE-III 2.
FT DOMAIN 216 303 FIBRONECTIN TYPE-III 3.
FT DOMAIN 304 392 FIBRONECTIN TYPE-III 4.
FT DOMAIN 393 470 FIBRONECTIN TYPE-III 5.
FT DOMAIN 471 562 FIBRONECTIN TYPE-III 6.
FT DOMAIN 563 652 FIBRONECTIN TYPE-III 7.
FT DOMAIN 653 741 FIBRONECTIN TYPE-III 8.
FT DOMAIN 742 830 FIBRONECTIN TYPE-III 9.
FT DOMAIN 831 921 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1150 1418 PROTEIN-TYROSINE PHOSPHATASE 1.
FT DOMAIN 1419 1711 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 1350 1350 BY SIMILARITY.

FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 982 982 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1711 AA; 187292 MW; 0F04D2D1A47A18A0 CRC64;

Query Match 3.4%; Score 96; DB 1; Length 1711;
Best Local Similarity 19.6%; Pred. No. 13;
Matches 107; Conservative 59; Mismatches 159; Indels 220; Gaps 31;

QY .131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCFDSIRTYTYADT--PDFFQLHNFSL 188
Db 519 DLVDLGPDTLSLTLSL-----VP-GSC-YTVSAWAWAGNLDSDSQKHSCTR 564

QY 189 PEEDTKLKIPIHRLALQIAQRPVSLASPTWTLKTNAGVNGKG----- 234
Db 565 PAPPTNL-----SLGFAHQPAALKAS-WYHPP-----GGDAFHRLRLPLTL 608
QY 235 -SLKGQPGDIYHOTWARYFVKFLDAYAEHLKQ---FWAVTAENEPSAGLLSGYPPQCLGF 290
Db 609 ESEKVLPREAQNFSAQ-----LTAGCFQVQLSTLWG---SERSSA-----NATGW 653
QY 291 TPEHQDFIARDLGPPTLANSTHNVRLMLDQRLLLPHWAKV-----VLTDPAAK 342
Db 654 TPTS-----APTLVNVTS-----DAPTQLQVSAHVPPGGRSRYQVTLQEESTR 696
QY 343 YVHGI-----AVHWYLDFLAPAKATLGETHRLFPNTMLFASE- 379
Db 697 TATSIMGPKEDGTSFLGLTPGTKYKVEVISWAGPLYTAANVSATYPLIPNELLVSMQA 756
QY 380 -----AC-----VGSKEFWEQSVRLGS---WDRGMQYSHSI----- 406
Db 757 GSAVNVNLAWPSPGLQGACHAQLSDAGHLSWEQPLKGOBLFMLRLDTPGHTISMSVRCR 816
QY 407 -----ITNLLYH-----VVGWT-----DWNLA-----LNPEGGPN 431
Db 817 AGPLQASTHLVLSVEPGVEDVLCHPEATYALNWTMPAGDQVCLVYVVERLVPGGGTH 876
QY 432 W---VRNEVDSPIIIVDTKDTFYKQPMFVHLGHFSKFIPEGSQRVGLVASQKNDLDAVAL 488
Db 877 FVFQVNTSGDALLPLNLMPTTSYRLSLTV-LGRNSRW-----SRAVSLVCS-----TSAEA 926
QY 489 MHP-----DGSVVVVLNRSKDVPLTIKDPVAV---GFLETIS-----PGYSI 528
Db 927 WHPPELAEPPOVELGTGMGTVMRGMFGKD-----DGOIQWYGIATINMTLAQPSREA 980
QY 529 HTYLW 533
Db 981 INYTW 985

RESULT 11
HM22_ARATH STANDARD; PRT; 512 AA.
AC 004921; 023623;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferrochelatase II, chloroplast precursor (EC 4.99.1.1) (Protoheme ferro-lyase) (Heme synthetase).
GN AT2G30390 OR T06B20.24 OR T09D09.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seedling;
RX MEDLINE=98426386; PubMed=9753778;
RA Chow K.-S., Singh D.P., Walker A., Smith A.G.;
RT "Two different genes encode ferrochelatase in Arabidopsis: mapping, expression and subcellular targeting of the precursor proteins."; Plant J. 15:531-541(1998).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana."; Nature 402:761-768(1999).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=97354168; PubMed=9210462;
RA Roper J.M., Smith A.G.;
RT "Molecular localisation of ferrochelatase in higher plant chloroplasts."; Eur. J. Biochem. 246:32-37(1997).
RL
CC -!- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX. MAY HAVE A ROLE IN DEALING WITH OXIDATIVE STRESS.
CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).
CC -!- PATHWAY: Protoheme biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST (THYLAKOID AND ENVELOPE MEMBRANES).
CC -!- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: U93215; AAB63095.1; -;
CC EMBL: Y13156; CAA73614.1; -;
CC HSP; P32396; IFFJ.
DR InterPro: IPR001015; Ferrochelatase.
DR Pfam: PF00762; Ferrochelatase; 1.
DR ProDom: PD002792; Ferrochelatase; 1.
DR TIGFAMS; TIGR00109; hemh; 1.
DR PROSITE; PS00534; FERROCHELATASE; 1.
KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 ?
FT CHAIN ? 512 CHLOROPLAST (POTENTIAL).
FT DOMAIN 9 19 FERROCHELATASE II.
FT DOMAIN 23 26 POLY-SER.
FT DOMAIN 80 83 POLY-PRO.
FT DOMAIN 90 93 POLY-SER.
FT CONFLICT 156 159 SKEG -> ARR (IN REF. 1).
SQ SEQUENCE 512 AA; 56618 MW; 521B1FC596E84A7 CRC64;
Query Match 3.3%; Score 95; DB 1; Length 512;
Best Local Similarity 18.9%; Pred No. 3;
Matches 102; Conservative 72; Mismatches 172; Indels 194; Gaps 26;
QY 67 PPTFPALGT-----FSRYESTRSGRRMELSMGPIQANHGTGLLTLPQEQ 112
Db 25 PPLLPQLSNDQSQRVVMHCTRLPFEAFATSSNRLLGKHSPLRA-----ALVTSNP-- 76
QY 113 KFQKVGFGGAMTDAALNILALSPPAQNLLLSYFSEEG-IGYNIIRV--PMASCDFSI 169
Db 77 -----LNISSSSVISDAISSSVITDDAKIGVLLNLGGP----- 111
QY 170 RYTYVADTPDDFQLHNFSLPEEDTKLKIPIHRLALQIAQRPVSLASPTWTLKTNGA 229
Db 112 -----ETLDDVQPFPLNLFADPDIIIRLPV---FQLQKPLAQFISVARAPKSKEGYAS 162
QY 230 VNGKSLKGQPGDIYHOTWARYFVKFLDAYAEH-KLQFWAVTAENEPSAGLLSGYPPQCL 288
Db 163 IGG-----GSP--LRHIT-----DAQAEELRKLCKEKNVPAKVYVGMRYWHPF--- 203
QY 289 GFTPEHQDFIARD-----LGPTLANSTH-HNVRL-----MLDDQRLLLPH 329
Db 204 ---TEEAIEQIKRGITKLVPLPLYPQFSISTSGSSRLRLERIFREDEYLVNMQHTVIPS 260
QY 330 W-----AKVLTDPAAKY-----VHGTAHVWYLDFLAPAKATLGE----- 365
Db 261 WYQREGYIKAMANLIQSELGKFGSPNQVWIFFSAHGVPPLAYVEEAGDPYKAEMECCVDLI 320
QY 366 ----THRLFPNTMLFASACVSGKEWQSVRLGSDRGMQYSHSIITNLLYHVVGWTDWN 421

Db 321 MEELDKRKITNAYTLAYQSRVGPVEWLK-----PYTEEAITEL----- 358
QY 422 LALNPEGGPNNVNFVDSPI-----IVDITKDTFYK-----Q 453
Db 359 -----GKKGVENLLAVPISFVSEHETLEEI--DVEYKELALKSGIKNGRVPALGTE 409
QY 454 PMFYHLGHFSKFIPEGSRQVLVASQKNDLDAVALMHPDGS--VVVVLNRSSKDVPLTI 511
Db 410 PMF--ISDLADAVVESLPYVGAMA--VSNLEARQSLVPLGSVELLATYDSQRRELPAV 465

RESULT 12
C11A_DASAM
ID C11A_DASAM STANDARD; PRT; 514 AA.
AC Q92045;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cytochrome P450 11A1, mitochondrial precursor (EC 1.14.15.6) (CYPX1A1)
DE (P450(SCC)) (Cholesterol side-chain cleavage enzyme) (Cholesterol
desmolase) (Fragment).
GN CYP11A1.
OS Dasyatis americana (Southern stingray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hypnosqualea; Pristiogaster; Batoidae;
OC Myliobatiformes; Myliobatoidei; Dasyatidae; Dasyatis.
OX NCBI_TaxID=50411;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Interranal gland;
RX MEDLINE=97225805; PubMed=9073075;
RA Nunez S., Trant J.M.;
RT "Isolation of the putative cDNA encoding cholesterol side chain
cleavage cytochrome P450 (CYP11A) of the southern stingray (Dasyatis
americana).";
RL Gene 187:123-129(1997).
CC -1- FUNCTION: CATALYZES THE SIDE-CHAIN CLEAVAGE REACTION OF
CHOLESTEROL TO PREGNENOLONE.
CC -1- CATALYTIC ACTIVITY: Cholesterol + reduced adrenal ferredoxin +
O(2) -> pregnenolone + 4-methylpentanal + oxidized adrenal
ferredoxin + H(2)O.
CC -1- PATHWAY: INITIAL RATE-LIMITING REACTION IN THE SYNTHESIS OF
VARIOUS STEROID HORMONES.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U63299; AAC60095.1; -.
HSSP; P00189; 1SCC.
InterPro; IPR001128; Cytochrome_P450.
Pfam; PF00067; p450; 1.
PROSITE; PS00086; CYTOCHROME_P450; 1.
Oxidoreductase; Monooxygenase; Membrane; Heme; Mitochondrion;
Steroidogenesis; Transit peptide.
NON_TER 1 1
TRANSIT <1 39 MITOCHONDRION (BY SIMILARITY).
CHAIN 40 514 CYTOCHROME P450 11A1.
BINDING 461 461 HEME (BY SIMILARITY).
SQ SEQUENCE 514 AA; 59863 MW; D0F03E25D5534FB2 CRC64;

Query Match 3.3%; Score 95; DB 1; Length 514;
Best Local Similarity 20.1%; Pred. No. 3;
Matches 92; Conservative 51; Mismatches 136; Indels 178; Gaps 24;
QY 168 SIRTYYADTPDDFQLHNFSLPEEDTKL-----KIPLIHALQLAQRVSLLA 215

Db 5 SILASTYA-----QRGSFTTPEHDTFLFPHRHSVTSESRIPEQTLKSLTDIP----- 53
QY 216 SPWTSPTWL-----KTNGAVNGKSLKGQPGDIYHOTWARYFVKFLDAYAEHKLQFWAV 269
Db 54 GNWRK-NWLVVYFWSNGLNNA-----HQWMLDNFNKYGPYREKIAYYESI 100
QY 270 TAENEPSAGLL--SGYFFQCLGFTPEH-----QRDFIARDLPTLANSTHHNVRLLM 319
Db 101 NIINPADAVIMNKSEGF-----PKRIEMAPWVAYRDLRKENYGVQLLNGENWKTRLI 154
QY 320 LDDQ-----RLLPHWAKVVL-----TDPEAAKYVHGIAVHWY- 352
Db 155 LNNISIFAQSSIQRLVPLENEVVDVFSVMVHKEVKSRSYWKTDLTNDLFKLALEVICYI 214
QY 353 -----LDFL-----APAKATLGETHRLFPNT--MLPASEA---CVGSKFWEQSVRLGSW 396
Db 215 LYGERLDLLQRYKYNKAPQK-FIDSIATMFHSTPIMLYVPPSLLKSINSKIWOQHV--GSW 271
QY 397 D-----RGMQYSHS---IITNLLYH----- 413
Db 272 DNIFEHADTYLKKAYRQFGGSKNEHAFPGVLFELLQALPFDIRASIIDVMGSAIDT 331
QY 414 ---VVGWTDWNLALNP-----EGPNWVRNFVDSPIIVDTKDTFYKQ 453
Db 332 TSFTVHMMYELAKHPHIQKNVRSEIMEAHQKTEGDP--VKMLKSVPLLKCVVKETLRLY 389
QY 454 PMFYHLGHF-----SKFIPEGs-QRVGLVASQKN 481
Db 390 PVAISIORYLNEEDTVLQNYHIPAGTLVQLGLYAMGRN 426

RESULT 13
GALC_HUMAN
ID GALC_HUMAN STANDARD; PRT; 669 AA.
AC P54803;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactocerebroside precursor (EC 3.2.1.46) (GALCERASE)
DE (Galactosylceramidase) (Galactosylceramide beta-galactosidase)
DE (Galactocerebroside beta-galactosidase).
GN GALC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta, and Skin fibroblast;
RX MEDLINE=94128088; PubMed=8297359;
RA Sakai N., Inui K., Fujii N., Fukushima H., Nishimoto J.,
Yanagihara I., Isegawa Y., Iwamatsu A., Okada S.;
RT "Krabbe disease: isolation and characterization of a full-length cDNA
for human galactocerebroside";
RL Biochem. Biophys. Res. Commun. 198:485-491(1994).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-59 AND 436-454.
RC TISSUE=Brain, and Testis;
RX MEDLINE=94108435; PubMed=8281145;
RA Chen Y.Q., Rafi M.A., de Galla G., Wenger D.A.;
RT "Cloning and expression of cDNA encoding human galactocerebroside,
the enzyme deficient in globoid cell leukodystrophy.";
RL Hum. Mol. Genet. 2:1841-1845(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324938; PubMed=7601472;
RA Luzi P., Rafi M.A., Wenger D.A.;
RT "Structure and organization of the human galactocerebroside (GALC)
gene.";
RL Genomics 26:407-409(1995).
RN [4]
RP SEQUENCE FROM N.A.

FT VARIANT 247 247 A -> T (IN GLD).
FT FTID=VAR_003388.
FT G -> S (IN GLD).
FT FTID=VAR_003389.
FT G -> D (IN GLD).
FT FTID=VAR_003390.
FT N -> T (IN GLD).
FT FTID=VAR_003391.
FT S -> F (IN GLD).
FT FTID=VAR_003392.
FT P -> A (IN GLD).
FT FTID=VAR_003393.
FT R -> W (IN GLD, BILATERAL CHERRY RED SPOTS).
FT FTID=VAR_003394.

Query Match 3.3%; Score 94.5; DB 1; Length 669;
Best Local Similarity 20.3%; Pred. No. 4.8;
Matches 87; Conservative 53; Mismatches 153; Indels 135; Gaps 24;

QY 119 GFGGAMTDAALNIALSP-PAQNLLKSYFSEE-GIGYNIIRVPMA---SCDFSIRTY 172
DB 41 GIGAVSGGATSRLLVNPYEPYRSQILDYLFKPNFGASHLKVEIGDGGQTTDGTPEPSH 100
QY 173 T-YADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRP-VSLLASPWTPTWLKTNGAV 230
DB 101 MHYALDENYFRGYEWLMKEAKK-----RNPNTLIGLPWSFPGWL----- 141
QY 231 NGKG----SLKGQPGDIYHQTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPFQ 286
DB 142 -GKGFDPVYNQLTAYVVVTV-----IVGAKRYHDLIDYIGIWNERS----- 184
QY 287 CLGFTPEHQRFIARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKV---VLTDPAAKY 343
DB 185 -----YNANYI-KILRKMLNYQGLQVRKIIASDNL-----WESISASMLLDALFKV 230
QY 344 VHGIADVHYLDLAPAKATLGETHRLFPNTMLFASEACVSKFW-----EQSVRLGS 395
DB 231 VDVIAGH-----YPGTHSAKDAKLTGKKLWSSSEDFSTLNSDMGAGC 271
QY 396 WDR--GMQYSHSITNLLYHVVGWTDNMLN-----PEG-----GPNWVRNF-VDSP 440
DB 272 WGRILNQYINGYMTSTI-----AWNLSVSYEQLPYGRCGLMTAQEPWSGHYVVEP 324
QY 441 IIVDITKOTFYKOPMFYHL---GHFSKEIPEGSQVRGLVASQKNLDLDAVALMHPDGSVV 497
DB 325 VWVS-AHTTQFTQGWYLYKTVGHLEK---GGSY-----VALTDGLGNLTI 366
QY 498 VVLNRSSK 505
DB 367 ILETMSHK 374

RESULT 14
PHK_LACLA STANDARD; PRT; 822 AA.
AC Q9CFH4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable phosphoketolase (EC 4.1.2.-).
GN LL1502.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403."

RL Genome Res. 11:731-753(2001).
CC -!- COFACTOR: Thiamine pyrophosphate (Potential).
CC -!- SIMILARITY: BELONGS TO THE XFP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE006381; AAK05600.1; -
DR InterPro; IPR000399; TPP_enzyme.
DR PROSITE; PS60002; PHOSPHOKETOLASE_1; 1.
DR PROSITE; PS60003; PHOSPHOKETOLASE_2; 1.
DR PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SQ SEQUENCE 822 AA; 93363 MW; C686D569D3E8E22F CRC64;

Query Match 3.3%; Score 94.5; DB 1; Length 822;
Best Local Similarity 20.5%; Pred. No. 6.4;
Matches 103; Conservative 61; Mismatches 180; Indels 159; Gaps 27;

QY 102 TGLLLTLQPEQKFKVKFGGAMT-----DAAALNIAL-----SPP- 138
DB 156 TGAILD-QPEQIAFAVVGDEAETGPLMTSWHSIKFINPKNDGAILPILDNGFKISNPT 214
QY 139 ----AQNLLKSYFSEEGIGYNIIRVPMSCDFSIPTYADTPDDFQLHNFSLPEEDTK 194
DB 215 LFARTSDVDIRKFF--EGLGYS-----PRYIENDDIHDYMAHYH-KLAAEVFD 258
QY 195 LKIPLIHRALQLAQR-----PVSLLASP--WTSPTWLKTNGAVNGKSLKGQP 240
DB 259 KAIEDIHQIQKDAREDNRYQNGEIPAWPIVIARLPKGGPRYNDWSGP---KFDGKGMP 315
QY 241 GDYHQTWARY-----FVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPF 285
DB 316 --IEHSFRAHQVPLPLSSKNMGTLPEFVKWMTSYQPETL-----F 353
QY 286 QCLGFTPEHQRFIARDLGPTLANSTHNNVRLMLDDQRLLLPHWAKV--LTDPEAAKY 343
DB 354 NADGSLKEELRDFAPKGMKMASNPVTNG---GVDSSNLVLPDQWQEFANPISNNRGKL 409
QY 344 VHGIADVHYLDLAPAKATLGETHRLFPNTM-LEASEACVSKFWEQ-SVRLGSW----- 396
DB 410 LPDTNDNMDMNVLSKYFA---EIVKLNPTFRFLRFGPDETMNSRFWEMFKVTNRQWQVIK 466
QY 397 ---DRGMQYSHSITNLL--YHVVGWTDNMLNPEGGP-----NW 432
DB 467 NPDEFISPEGRIIDSQLEHQAEGWLE-GYTLTGRTGAFASYESFLRVVDSMLTQHFKW 525
QY 433 VRNEVDS-----PIIVDITKDTFYKOPMFYHLGHFSKFIPEGSQVRGL---VASQKNDL 483
DB 526 IRQAADQKWRHDYPSLNVISTSTVFQQD---HNGY-----THQDPGMLTHLAEKKKSD 575
QY 484 DAVALMHPDGSVAVVVLNRSSK 506
DB 576 -IRQYLPADGNTLLAVFDRAFQD 597

RESULT 15
TYCC_BACBR STANDARD; PRT; 6486 AA.
ID TYCC_BACBR
AC O30409;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III [Includes: ATP-dependent asparagine
DE adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine
DE adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE adenylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylase
DE (ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA)]

DE (Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine
DE activase)).
GN TYCC.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D.; Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylation domains."
RL J. Bacteriol. 179:6843-6850(1997).
CC -!- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
CC GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC PEPTIDE PRODUCT.
CC -!- COFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -!- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.
CC -!- DOMAIN: CONSISTS OF SIX MODULES, AND HARBOURS A PUTATIVE
CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).
CC -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF004835; AAC45930.1; -
HSSP; P14687; 1AMU.
InterPro; IPR000873; AMP-bind.
InterPro; IPR001242; Condensatn.
InterPro; IPR003880; Ppantne_attach.
InterPro; IPR000379; Ser_estrs_site.
InterPro; IPR001031; Thioesterase.
Pfam; PF00501; AMP-binding; 6.
Pfam; PF00550; pp-binding; 6.
Pfam; PF00668; Condensation; 6.
Pfam; PF00975; Thioesterase; 1.
PRINTS; PR00154; AMPBINDING.
PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
PROSITE; PS00455; AMP_BINDING; 6.
PROSITE; PS50075; ACP_DOMAIN; 6.
Ligase; Antibiotic biosynthesis; Phosphopantetheine;
Multifunctional enzyme; Repeat.
REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.

| | | | | |
|----|----------|------|--|-------------------------------------|
| FT | DOMAIN | 5124 | 5191 | ACYL CARRIER (ACP) 5. |
| FT | DOMAIN | 6167 | 6234 | ACYL CARRIER (ACP) 6. |
| FT | BINDING | 1000 | 1000 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| FT | BINDING | 2037 | 2037 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| FT | BINDING | 3075 | 3075 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| FT | BINDING | 4110 | 4110 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| FT | BINDING | 5154 | 5154 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| FT | BINDING | 6197 | 6197 | PHOSPHOPANTETHEINE (BY SIMILARITY). |
| SQ | SEQUENCE | 6486 | AA; 724011 MW; 4934900AF07DF786 CRC64; | |

Query Match 3.2%; Score 92.5; DB 1; Length 6486;
Best Local Similarity 19.0%; Pred. No. 1.8e+02;
Matches 80; Conservative 54; Mismatches 127; Indels 161; Gaps 19;

| | | | |
|----|------|--|------|
| QY | 173 | TYADTPDDFQLHN--FSLPEEDTKLKIPLIHRALQLA----- | 207 |
| Db | 4686 | TFVPIDPDYPLERQAFLMLEDSEAKLLLTQLKMSQVAFPYETFYLDTEYDQDEETGNLEH | 4745 |
| QY | 208 | -QRPVSLASPTWTLKTNAGAVNGKSLKGGPDDIYHQTWARYFVKFLDAYAEHKLQF | 266 |
| Db | 4746 | VAQPNVAYIIYTSGTGKPKGVV-----IEHRSYANVAFAWKDEY--H----- | 4787 |
| QY | 267 | WAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPTLANSTHNVRLMLDDQRL | 326 |
| Db | 4788 | -----LDSPVRL-----QMAFADV-----STGDFARALLTGQQLVI | 4822 |
| QY | 327 | LPHWAKVLTDPBAKYVHGIAVHWYLDLAPAKATLGETHRLFPN----- | 372 |
| Db | 4823 | CPNGVKM--DP--ASLYETIRREITIFEATPALIMPLMHYVYENELDMSQMKLLILGA | 4877 |
| QY | 373 | -----TMLF-----ASEACVGSKFWEQ-----SVRLGSWDRGMQYSHSI | 406 |
| Db | 4878 | DSCPAEDFKTLARFGQKMRINSYGVTEACIDTSYEEETDVTAIRSGTVPIGKP---- | 4932 |
| QY | 407 | ITNLLYHVVGWTDNWLALNPE-----GGPNWVRNFVDSPIIVDITKDTFYKQPM---- | 455 |
| Db | 4933 | LPNMTMYV---DAHLNLQPVGVVVGELCIGGAGVARGYLNRP---ELTEKFVNPFPAPG | 4986 |
| QY | 456 | --FYHLGHFSKFIPEGSQ-----RVGLVASQKNDLDAVALMHPDGSVAV | 496 |
| Db | 4987 | ERLYRTGDLAKWRADGNVEFLGRNDHQVKIRGVRIELGEIETQLRKLDGIT-----EAV | 5040 |
| QY | 497 | VV 498 | |
| Db | 5041 | VV 5042 | |

Search completed: February 20, 2003, 11:18:37

Job time : 22 secs

